

UNIVERSIDAD CARLOS III DE MADRID

ESCUELA POLITÉCNICA SUPERIOR

INGENIERÍA TÉCNICA DE INFORMÁTICA DE GESTIÓN



PROYECTO FINAL DE CARRERA

# Rectificación de Imágenes

AUTOR: Oscar Pérez Alfonso

TUTOR: Arturo de la Escalera Hueso

Leganés, 21 de Diciembre del 2010

# Índice general

<b>1. Introducción</b>	<b>7</b>
1.1. Planteamiento del problema . . . . .	8
1.2. Estructura del documento . . . . .	9
<b>2. Detección de características.</b>	<b>10</b>
2.1. Detección por Harris . . . . .	11
2.2. Detección SURF . . . . .	16
2.2.1. Detector Hessiano . . . . .	16
2.2.2. Descriptor SURF . . . . .	17
2.2.2.1. Asignación de Orientación . . . . .	18
2.2.2.2. Componentes del Descriptor . . . . .	18
2.2.3. Invariancia Detectores . . . . .	19
<b>3. Geometría</b>	<b>21</b>
3.1. Modelo de cámara . . . . .	21
3.1.1. Geometría Proyectiva y Coordenadas Homogéneas . . . . .	21
3.1.2. Cámara Estenopeica . . . . .	22
3.1.3. Distorsión de las lentes . . . . .	24
3.1.4. Calibración . . . . .	25
3.2. Introducción a la Geometría Epipolar . . . . .	27
3.2.1. Matriz esencial . . . . .	29
3.2.2. Calibración de un Sistema Estéreo . . . . .	30
3.2.3. Matriz fundamental. . . . .	30

## *Índice general*

3.3. Rectificación de imágenes . . . . .	34
3.3.1. Rectificación de Hartley . . . . .	36
3.3.2. Rectificación de Whelan . . . . .	39
<b>4. Resultados</b>	<b>43</b>
4.1. Medición de la calidad de las correspondencias . . . . .	43
4.1.1. Definición error de reproyección . . . . .	44
4.1.2. Comentarios . . . . .	45
4.2. Medición de las distorsiones provocadas por la rectificación . . . . .	47
4.2.1. Comentarios . . . . .	51
4.3. Proceso Completo . . . . .	55
<b>5. Conclusiones</b>	<b>58</b>
<b>A. Apéndices</b>	<b>60</b>
A.1. Presupuesto . . . . .	60
A.2. Imagen integral . . . . .	62
A.3. Homografía . . . . .	63
A.4. Deducción de homografía pareja en Hartley . . . . .	64
A.5. Método Nelder-Mead simplex . . . . .	66
A.6. Imágenes de Correspondencias . . . . .	67
A.7. Imágenes de Rectificaciones . . . . .	71
<b>Bibliografía</b>	<b>78</b>

# Índice de figuras

1.0.1.Posibles aplicaciones de la visión artificial. . . . .	8
2.1.1.La misma imagen a la que le aplicamos la detección de Harris con entornos de 3 píxeles (a), y de 11 píxeles (b). . . . .	12
2.1.2.Maya de gradientes sobre un cuadrado. . . . .	14
2.1.3.Piramide gaussiana de una imagen. . . . .	14
2.1.4.Parches orientados basados en la detección de Harris. . . . .	15
2.2.1.Aproximación de derivadas gaussianas. . . . .	17
2.2.2.La marcación de “gotas” que selecciona el detector de SURF. . . . .	18
2.2.3.Basados en la respuesta de Haar-wavelets se puede estimar la dirección en la que ocurren cambios en la intensidad de la imagen.. . . .	19
3.1.1.Modelo cámara estenopeica . . . . .	23
3.1.2.Modelo Proyectivo de cámara esteponeica . . . . .	24
3.1.3.Distorsión radial generada por la distinta inclinación con la que los rayos atraviesas la lente en diferentes zonas. . . . .	25
3.1.4.Distorsión tangencial producida por la inclinación del sensor respecto de la lente. . . .	25
3.1.5.Efectos de las distorsión radial y tangencial en una imagen real modificada artificialmente.	26
3.1.6.Calibración una cámara capturando un objeto conocido en diferentes posiciones. . . . .	26
3.2.1.Sistema de cámaras estéreo y su geometría epipolar. . . . .	28
3.2.2.Posibles hipótesis para un punto basadas en la línea epipolar . . . . .	28
3.2.3.La matriz esencial contiene información sobre la rotación y la traslación entre las cámaras.	29



## *Índice de figuras*

3.2.4.La transferencia de puntos entre las dos imágenes a través de un plano es la matriz fundamental. . . . .	30
3.2.5.Imprecisión de la matriz fundamental. . . . .	34
3.3.1.El objetivo de la rectificación es alinear las filas de las 2 imágenes. . . . .	35
3.3.2.Torsión introducida en la imagen izquierda en la rectificación. . . . .	38
4.1.1.Evaluación en imágenes de las correspondencias obtenidas con Harris+CN de la escena Drive. . . . .	46
4.1.2.Evaluación en imágenes de las correspondencias obtenidas con Harris+CNME de la escena Arch. . . . .	46
4.1.3.Evaluación de las correspondencias en la escena Arch con SURF, en verde aceptadas, en rojo rechazadas. . . . .	48
4.2.1.Resultado de aplicar diferentes métodos de rectificación a la pareja de imágenes original y con las líneas epipolares (a) y (b), de la escena Roof. (c) y (d) rectificación de Hartley , (e) y (f) Hartley con corrección de torsión, (g) y (h) rectificación de Whelan. . . . .	52
4.2.2.Fragilidad de la matriz fundamental provocando diferentes imágenes rectificadas con el método de Whelan. . . . .	54
4.3.1.Pareja de imágenes originales. . . . .	55
4.3.2.Detección de características SURF. . . . .	55
4.3.3.Selección de equivalencias entre imágenes. . . . .	56
4.3.4.Cálculo de la matriz fundamental . . . . .	56
4.3.5.Rectificación de imágenes. . . . .	57
A.2.1Imagen integral . . . . .	62
A.3.1Elección sistema de referencia para el cálculo de la homografía que los relaciona. . . .	64
A.5.1Fases del método simplex de optimización, a) reflexión, b) expansión, c) contracción, 2 casos ,posibles, d) reducción. . . . .	66
A.6.1Imágenes de correspondencias de Harris+CN para (a y b) escena Arch, (c y d) escena Drive, (e y f) escena Lab, (g y h) escena Roof. . . . .	68

## *Índice de figuras*

A.6.2	Imágenes de correspondencias de Harris+CNME para (a y b) escena Arch, (c y d) escena Drive, (e y f) escena Lab, (g y h) escena Roof. . . . .	69
A.6.3	Imágenes de correspondencias de SURF para (a y b) escena Arch, (c y d) escena Drive, (e y f) escena Lab, (g y h) escena Roof. . . . .	70
A.7.1	Rectificación de imágenes para la escena Arch, primera fila, original y original con líneas epipolares, segunda fila Hartley, tercera fila, Hartley + Zhang, cuarta fila Whelan . . .	72
A.7.2	Rectificación de imágenes para la escena Drive, primera fila, original y original con líneas epipolares, segunda fila Hartley, tercera fila, Hartley + Zhang, cuarta fila Whelan. . .	73
A.7.3	Rectificación de imágenes para la escena Lab, primera fila, original y original con líneas epipolares, segunda fila Hartley, tercera fila, Hartley + Zhang, cuarta fila Whelan . . .	74
A.7.4	Rectificación de imágenes para la escena Roof, primera fila, original y original con líneas epipolares, segunda fila Hartley, tercera fila, Hartley + Zhang, cuarta fila Whelan . . .	75
A.7.5	Rectificación de imágenes para la escena Slate, primera fila, original y original con líneas epipolares, segunda fila Hartley, tercera fila, Hartley + Zhang, cuarta fila Whelan . . .	76
A.7.6	Rectificación de imágenes para la escena Yard, primera fila, original y original con líneas epipolares, segunda fila Hartley, tercera fila, Hartley + Zhang, cuarta fila Whelan . . .	77

# 1. Introducción

Se puede decir que la visión artificial es un campo con 30 años de historia aproximadamente, a pesar de esto, queda un largo camino por recorrer encontrándose en pleno desarrollo. Con aplicaciones en prácticamente cualquier actividad, robótica, vehículos autónomos, automatización industrial, inspección de procesos, de edificaciones, vigilancia, industria gráfica, medicina..., etc, las líneas de investigación son muchas.

Dentro de la automática, el poder dotar a los robots de algún tipo de visión, aunque sea a niveles básicos, puede abrir numerosas posibilidades en las automatización de diversos procesos. Conseguir deducir el entorno a través de imágenes es aun largo, sin embargo, en los últimos años ha habido una sucesión de logros importantes debido al aumento exponencial de la capacidad de computación, al abaratamiento de las cámaras de vídeo, y al desarrollo de técnicas que parecen haber encontrado el buen camino.

Pueden identificarse dos acercamientos distintos para percibir el entorno a partir de imágenes, SFM, Structure From Motion, y SLAM, Simultaneous Localization And Mapping, cada técnica es una simplificación del problema de conocer la estructura del entorno y conocer la relación con él. SFM deduce la estructura limitando el movimiento de la cámara, por ejemplo, deducir la estructura de un objeto manteniendo la cámara fija y girando el objeto, o grabar un edificio limitando el posible desplazamiento de la cámara con el transcurso del tiempo. SLAM intenta detectar el movimiento de la cámara, su localización en cada momento, estableciendo restricciones sobre la estructura del entorno, utilizando cámaras de gran angular y “High-Speed” para hacer al entorno más estable, con heurísticas que limitan las posibles localizaciones.

En la raíz de todos estos procesos nos encontramos con la pregunta de ¿ cómo reconocer partes de una imagen en otra ? . El problema de correspondencia entre imágenes es la base para poder inferir la estructura del entorno deduciendo las relaciones geométricas que las imágenes tienen entre si.

## 1. Introducción

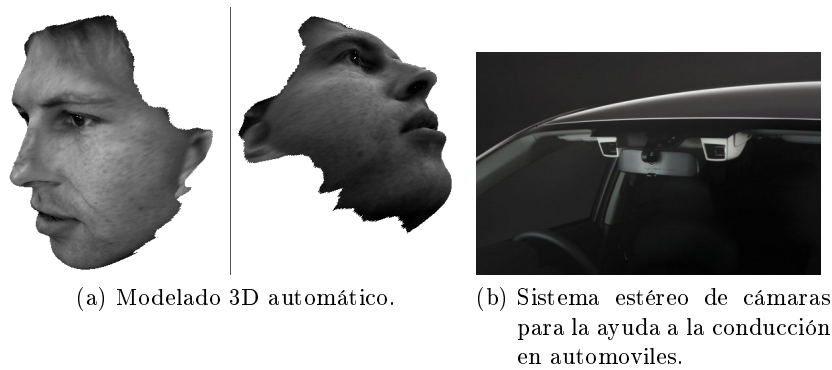


Figura 1.0.1.: Posibles aplicaciones de la visión artificial.

### 1.1. Planteamiento del problema

Para conocer la distancia a la que se encuentra un punto podemos tomar imágenes de este con un sistema estéreo de cámaras, el punto será proyectado en cada imagen por lo que conociendo la relación entre las imágenes podremos deducir la distancia. El atractivo de este sistema es que obtenemos una estimación de la profundidad de los objetos que perciben las cámaras en un instante de tiempo, como lo haría un láser, con las ventajas de las cámaras, menor coste y mayor frecuencia de muestreo. Esta capacidad facilita la tarea del seguimiento de objetos y a su vez da mayor confianza en la composición de esta información en un mapa 3D del entorno, siendo superior sobre los sistemas monoculares que se ven limitados por esquemas probabilistas sobre el movimiento de las imágenes y sus transformaciones en 2D.

Para sustentar este proceso, el problema de correspondencia entre el par de imágenes es el primer paso a realizar, y lo restringimos a la relación entre las imágenes obtenidas por dos cámaras en posiciones similares, aunque el enunciado del problema parece sencillo veremos que no es en absoluto trivial.

Vamos a estudiar diferentes posibilidades que hay para obtener correspondencias entre puntos de las imágenes modificándolas de tal forma que limita las posibles correspondencias entre las imágenes, y por tanto reduce el espacio de búsqueda de correspondencias correctas.

## **1.2. Estructura del documento**

En el capítulo 2 veremos diferentes posibilidades para la detección de características en las imágenes, y como describir porciones de la imagen para así poder hacer comparaciones y hacer hipótesis de correspondencias entre 2 imágenes.

En el capítulo 3 veremos cuales son las relaciones geométricas que hay en 1 par de imágenes, que es lo que nos pueden decir, como se puede reducir las posibilidades de correspondencias entre las imágenes y como se calculan.

En el capítulo 4 veremos resultados de la implementación de los dos capítulos anteriores en varias parejas de imágenes pertenecientes a 6 escenas diferentes.

## 2. Detección de características.

La detección de características en una imagen es la base fundamental de la resolución del problema de correspondencia entre imágenes. El ataque naif de este problema es comparar los píxeles de cada imagen entre si, teniendo  $N^2$  comparaciones, este computo no es asumible para computadoras estándar actuales si queremos un sistema operando en tiempo real. Se hace evidente la necesidad de emplear un método más eficaz, la pre-selección de píxeles por diferentes criterios permite resolver el problema de correspondencia.

No todas las regiones de una imagen contienen la misma información, son interesantes aquellas que maximicen alguna función local sobre alguna región de la imagen que sea rápidamente evaluable, a estas regiones las denominados características. Hay que considerar que la información visual que proporciona una cámara puede ser altamente variable para una misma escena, cambios en iluminación, cambios en las “texturas visuales” de los objetos, cambios en formas y objetos al mover la cámara, oclusiones entre objetos, y un largo etc.

La bibliografía relacionada es muy extensa, si nos centramos en detectar una característica en una imagen e identificarla de nuevo en otra, podemos resumir que una característica debe ser:

- Fácilmente detectable, se maximiza claramente en su entorno
- Robusta, pequeños cambios locales la afectan levemente
- Identificable, maximiza la distancia respecto de alguna función de comparación

Pueden considerarse varios tipos de características según la entidad geométrica que represente, esquinas (puntos), líneas, bordes (curvas), regiones en general, se elige la selección de puntos ya que nos permite manejar la geometría que relaciona las imágenes de manera precisa. He utilizado la palabra punto porque entre las posibilidades que vamos a estudiar, aunque el resultado de la detección de

## 2. Detección de características.

características es en ambos casos puntos, la heurística para la selección de esos puntos difiere en su naturaleza.

Una de las propiedades deseable de una característica es que pudiera ser identificable, fácilmente distinguible de otras, por lo que a un punto hay que dotarle de atributos basados en su entorno para poder caracterizarlos, si puede ser, de manera única. El descriptor de un punto esta íntimamente relacionado con la manera en que se selecciono, por esta razón, en cada método se explica la detección de las esquinas seguido de la descripción que se hace de ellas.

Los métodos elegidos para la extracción de características son el método SURF, Speeded Up Robust Features, y el método de Harris.

Harris [2] lleva siendo durante mucho tiempo el método más utilizado en la detección de esquinas por su equilibrio entre rapidez y robustez, recientemente han surgido otras propuestas pero casi siempre basadas en la matriz hessiana. Estudiando los diferentes métodos de detección a través de comparativas publicadas [6, 7] se observa que:

1. Los detectores basados en la hessiana son más estables y distinguibles que los detectores basados solo en su traza como es Harris ya que permite distinguir mejor si son estructuras alargadas y/o mal localizadas.
2. Las aproximaciones como la diferencia de gaussianas puede proporcionar mayor rapidez a cambio de una pequeña perdida en la precisión de los resultados.

Respecto de los descriptores hay una gran variedad de posibilidades como filtros de derivadas gaussianas, filtros orientados, invariantes afines de textura, momentos invariantes, características locales basadas en su fase, etc. La utilización de descriptores representando la distribución de gradientes en pequeña escala en la vecindad de una esquina ha destacado entre las demás, mostrando los mejores resultados, esta opción fue introducida por Lowe en SIFT [8] y es en la que esta inspirado SURF.

### 2.1. Detección por Harris

Para la detección de las esquinas primero se buscan puntos interesantes que serán aquellos cuyas derivadas en un entorno alrededor del punto tienen características interesantes. Para la detección de puntos interesantes se calcula la matriz de covarianza en el entorno de cada punto de la matriz hessiana

## 2. Detección de características.

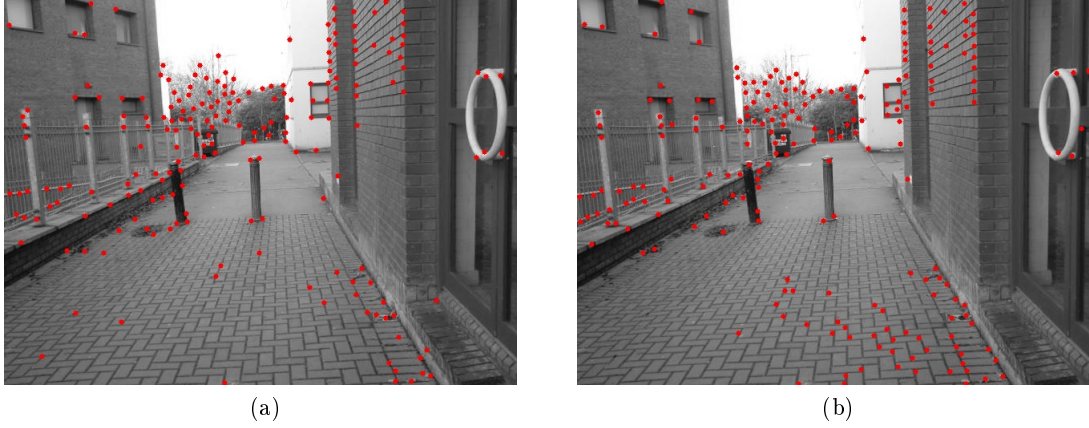


Figura 2.1.1.: La misma imagen a la que le aplicamos la detección de Harris con entornos de 3 píxeles (a), y de 11 píxeles (b).

de la imagen, el entorno de la imagen es simplemente una región cuadrada alrededor del punto, 3 o 5 píxeles es lo más utilizado, este parámetro puede influir considerablemente en las esquinas detectadas, tanto en su número como en su localización.

La matriz hessiana  $H(p)$  sobre un punto  $p = (x, y)$  se define como :

$$H(p) = \begin{bmatrix} \frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 I}{\partial x \partial y} \\ \frac{\partial^2 I}{\partial x \partial y} & \frac{\partial^2 I}{\partial y^2} \end{bmatrix}$$

Harris considera la auto-correlación (convolución en funciones continuas) de la matriz hessiana en la vecindad de un punto  $p$ .

$$M = \begin{bmatrix} \sum_{-k \leq i, j \leq k} w_{i,j} I_x^2(x+i, y+j) & \sum_{-k \leq i, j \leq k} w_{i,j} I_x(x+i, y+j) I_y(x+i, y+j) \\ \sum_{-k \leq i, j \leq k} w_{i,j} I_x(x+i, y+j) I_y(x+i, y+j) & \sum_{-k \leq i, j \leq k} w_{i,j} I_y^2(x+i, y+j) \end{bmatrix}$$

Donde  $w_{i,j}$  es una función de ponderación que suele ser una gaussiana.

Los autovalores de  $M$  nos permite clasificar los puntos:

- Si ambos autovalores son bajos, la función local de auto-correlación sera “llana” ( hay poco cambio en ese entorno) luego el cuadrado local seleccionado tiene una intensidad similar en todos sus puntos.



## 2. Detección de características.

- Si uno de los autovalores es alto y el otro bajo, la función local de auto-correlación tendrá la forma de un pico alargado, luego desplazamientos en una dirección (en la del pico) genera poco cambio, por contra en la dirección perpendicular hay un gran cambio, por lo tanto tenemos una línea-borde.
- Si ambos autovalores son altos, la función local de auto-correlación tendrá un pico bien definido, por lo que desplazamientos en cualquier dirección conlleva un gran cambio en la intensidad de la imagen, entonces tenemos una esquina.

Con la intención de acelerar el proceso Harris define la función:

$$det(M) - k \cdot traza(M)^2$$

Donde k es factor de Harris ajustable, de este modo se pueden detectar las esquinas buscando máximos locales sobre la matriz resultante de aplicar el operador de Harris.

Una vez que detectadas las esquinas en cada imagen se considera la opción que a priori parece va a dar mejores resultados, la correlación entre los “parches” extraídos con centro en cada esquina, con la que se espera una alta tasa de identificaciones correctas.

Para alcanzar invariancia sobre rotaciones seleccionamos una orientación para el parche alineándolos antes de calcular su correlación, para esto se divide la ventana del parche en pequeñas cuadrados con los que se crea el histograma de gradientes orientados. La magnitud y orientación en cada cuadrado se define como:

$$orientación(x, y) = \tan^{-1}\left(\frac{Sum_y}{Sum_x}\right)$$

$$magnitud(x, y) = \sqrt{Sum_x^2 + Sum_y^2}$$

Donde :

$$Sum_x(x, y) = \sum_{-k \leq i, j \leq k} I_x(x + i, y + j)$$

$$Sum_y(x, y) = \sum_{-k \leq i, j \leq k} I_y(x + i, y + j)$$

Después de hacer un suavizado con una gaussiana sobre los resultados de la magnitud, se elige la orientación dominante con una ventana deslizante de  $\Pi/3$  al igual que se hace en SURF.

## 2. Detección de características.

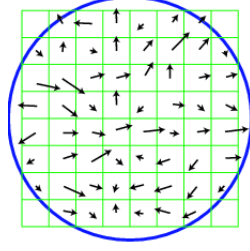


Figura 2.1.2.: Mapa de gradientes sobre un cuadrado.



Figura 2.1.3.: Pirámide gaussiana de una imagen.

Para calcular la correlación pueden elegirse diferentes formulas siendo deseable invariancia en las posibles diferencias entre el rango de percepción de luminosidad entre las cámaras, por esto se utiliza la correlación normalizada. Dados dos parches  $p_1$  y  $p_2$  ambos de dimensiones  $n \times m$  tendremos la media  $\mu$  y la varianza  $\sigma^2$  para cada parche, por ejemplo para  $p_1$  :

$$\mu_1 = \frac{1}{n m} \sum_{i=1}^n \sum_{j=1}^m p_1[i, j] \quad ; \quad \sigma_1^2 = \frac{1}{n m} \sum_{i=1}^n \sum_{j=1}^m (p_1[i, j] - \mu_1)^2$$

La correlación normalizada se define como :

$$cn = \frac{\sum_{i=1}^n \sum_{j=1}^m (p_1[i, j] - \mu_1)(p_2[i, j] - \mu_2)}{n m \sigma_1 \sigma_2}$$

Más adelante se vera que los resultados obtenidos son malos, es por esto que se probó una mejora aplicando este proceso sobre múltiple escalas.

## 2. Detección de características.



(a) En la escala original



(b) Multi Escala, la escala esta representada por el tamaño de los parches, separados por claridad.

Figura 2.1.4.: Parches orientados basados en la detección de Harris.

### Multi Escala

Se realizan pirámides gaussianas de la imagen a tratar ( 2.1.3), consiste en aplicar un suavizado de la imagen mediante la convolución con una gaussiana 2D reduciendo además la frecuencia de muestreo sobre la imagen, normalmente a la mitad, cogiendo un píxel y desechando otro intercaladamente en filas y columnas. La comparación de los parches no se modifica, todos tendrán el mismo tamaño aunque en realidad pertenezcan a dominios de espacio diferentes.

En la figura 2.1.4 se puede ver algunos ejemplos de la localización de los parches, en las imágenes de multi escala, se puede apreciar como en algunos puntos el centro de las cajas no parece corresponder con una esquina, esto no es un error, hay que tener en cuenta el sub-muestreo que se hace sobre la imagen. Esto puede suponer una dificultad para los cálculos posteriores de la geometría epipolar por lo que sería necesario solucionarlo en el futuro.

## 2.2. Detección SURF

El método de SURF, Speeded Up Robust Features [3, 4], es una variación del método SIFT, Scale Invariant Feature Transform, este ultimo es más pesado en términos de coste de computación y además su uso esta sujeto a patente, por lo que SURF surgió como una ayuda a la investigación en este campo estando libre de patentes y además manteniendo las características de repetibilidad (detectabilidad y distinguibilidad) de las esquinas SIFT pero con un menor coste en su cálculo.

Una debilidad de la detección de esquinas de Harris es que no es invariante a escalados, esto quiere decir que un objeto entre dos imágenes que tenga un cambio de tamaño no necesariamente muy grande, en la comparación de sus descriptores, seguramente no tendrán una respuesta demasiado alta. Lindeberg introdujo el concepto de la selección automática de escala lo que hace variable el entorno considerado en la descripción de esquinas, la aplicación de este concepto en diferentes métodos de descripción de esquinas ha incrementado el acierto en general de las hipótesis de correspondencias. Entonces surge la necesidad de una localización precisa de las esquinas que se realiza mediante la hessiana, adaptada a la escala seleccionada, o mediante el operador de Harris sobre esa hessiana. Posteriormente Lowe, en SIFT, aproxima la Laplaciana Gaussiana mediante el filtro de Diferencia de Gaussianas que son aproximadamente iguales. La detección en SURF tiene un enfoque similar, se basa en la matriz hessiana, pero es un detector laplaciano muy básico, utiliza filtros discretizados que se calculan muy rápidamente en las imágenes integrales [1] para reducir el tiempo de computación.

Para las aplicaciones, los tres pasos para establecer correspondencias son detección, descripción, comparación (emparejamiento), se deben intentar acelerar al máximo. En la fase de comparación, el gran numero de dimensiones de SIFT, vector de 128, 8 subintervalos por cada 4x4 cajas, es un factor determinante en su lentitud, por otra parte, la esquina se describe mediante la respuesta de Haar-wavelets, ondas Haar, en la vecindad del punto, y se utilizan 64 dimensiones.

### 2.2.1. Detector Hessiano

Dado un punto  $p = (x, y)$  en la imagen  $I$ , la matriz hessiana  $H(p, \sigma)$  en  $p$  a escala  $\sigma$  se define como :

$$H(p, \sigma) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix}$$

## 2. Detección de características.

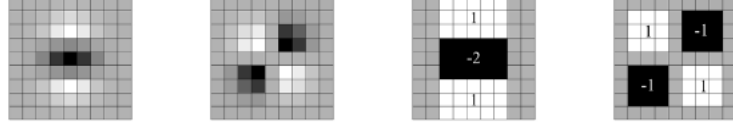


Figura 2.2.1.: Aproximación de derivadas gaussianas.

Donde  $L_{xx}(x, \sigma)$  es la convolución de la derivada gaussiana de segundo orden  $(dI/dx)^2$  con la imagen  $I$  en el punto  $x$ , y de manera similar para  $L_{xy}(x, \sigma)$  y  $L_{yy}(x, \sigma)$ . Aquí se realiza una aproximación con la discretización de las segundas derivadas gaussianas por medio de filtros del estilo de Viola-Jones que a su vez se calculan rápidamente gracias a las imágenes integrales (ver A.2) acelerando su procesamiento. Para garantizar la norma de Frobenius para cualquier tamaño de filtro-ventana-entorno se efectúa una normalización respecto del tamaño de la ventana.

El filtrado sobre diferentes escalas en el dominio de la imagen suele realizarse con pirámides gaussianas y sub-muestreo, gracias a los filtros de caja y a las imágenes integrales puede realizarse este proceso con la misma velocidad para todas las escalas simplemente aumentando el tamaño de los filtros y evaluando sobre la imagen integral.

Para localizar los puntos interesantes en la imagen en las diferentes escalas se realiza una detección de máximo locales sobre un entorno 3x3x3 sobre el que después se calcula la posición e interpola su valor en ese entorno ajustando una curva cuadrática basada en el entorno, para más detalles mirar[8].

La detección de características SURF se basa en la respuesta de un banco de filtros, de derivadas gaussianas, en diferentes escalas sobre la matriz hessiana, por tanto, a pesar de que su resultado es un punto, se puede considerar que detectan regiones con una determinada estructura (que necesariamente será similar al filtro aplicado). En la figura 2.2.2 se pueden apreciar lo que comentamos.

### 2.2.2. Descriptor SURF

El descriptor se procesa en 2 etapas, en la primera se calcula una orientación sobre una región circular alrededor del punto, en la segunda alineamos una región cuadrada con esa orientación para extraer el descriptor.

## 2. Detección de características.

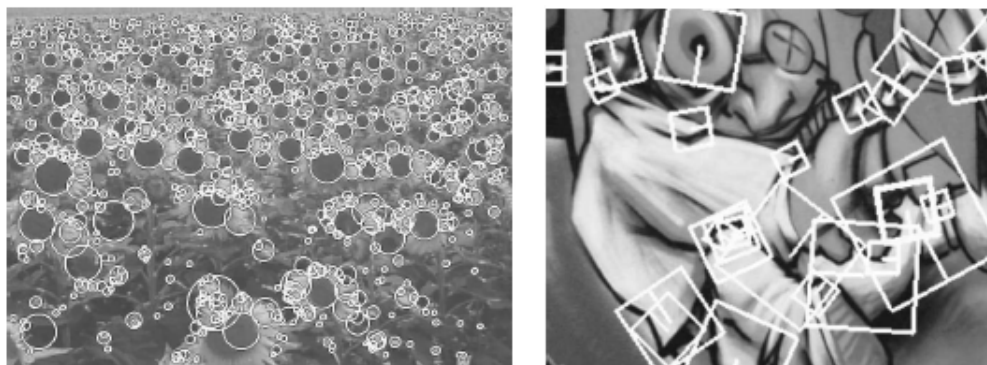


Figura 2.2.2.: La marcación de “gotas” que selecciona el detector de SURF.

### 2.2.2.1. Asignación de Orientación

Para conseguir la invariancia sobre rotación se asigna a cada punto una orientación en base a la información que hay en una región circular alrededor del punto. Con las respuestas de los filtros Haar-wavelet tenemos sus respuestas sobre los ejes X e Y que pueden traducirse como vectores en un eje de coordenadas. La orientación se selecciona sumando todas las respuestas en una ventana deslizante con un tamaño de cubriendo  $\Pi/3$  radianes, tamaño elegido experimentalmente. Un tamaño menor no es discriminatorio con las respuestas locales, y un tamaño mayor proporciona una orientación representativa de la región.

### 2.2.2.2. Componentes del Descriptor

Se construye una región cuadrada centrada sobre el punto y alineada con la orientación calculada en la etapa anterior. Esta región es dividida uniformemente en  $4 \times 4$  sub-cuadrados con lo que se captura la información espacial alrededor del punto. Para cada región se calculan las respuestas para los filtros con lo que se crea un vector para cada región con 4 componentes, la respuesta sobre X, la respuesta sobre Y, y los valores absolutos de los dos anteriores. Como teníamos 16 sub-cuadrados, tenemos un descriptor de 64 componentes. La respuesta de los filtros es invariable a cambios de iluminación y a cambios en el contraste.

Para un proceso de comparación más rápido, el signo del laplaciano de la traza de la matriz hessiana es utilizado para distinguir rápidamente regiones claras sobre fondos oscuros de lo inverso, regiones oscuras sobre fondos claros.

## 2. Detección de características.

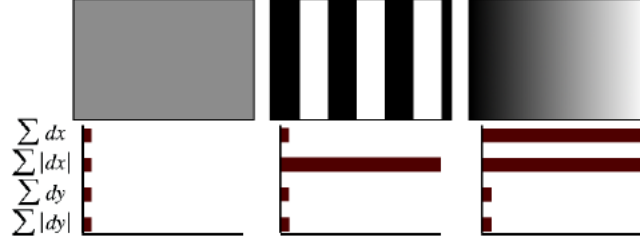


Figura 2.2.3.: Basados en la respuesta de Haar-wavelets se puede estimar la dirección en la que ocurren cambios en la intensidad de la imagen..

### 2.2.3. Invariancia Detectores

Descritos los dos métodos vamos a ver que su eficacia va a depender en gran medida de la capacidad de detectar los puntos interesantes soportando las diferentes variaciones en las que pueden presentarse las imágenes. Ambos métodos se utilizan sobre las imágenes en escala de grises, y los dos soportan alteraciones en el dominio de la intensidad. Las alteraciones geométricas en 2D pueden ser:

Isometría: Se mantienen las distancias entre todos los puntos, representa el movimiento que puede hacer sobre el plano de un objeto rígido.

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & t_x \\ \sin\theta & \cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

Similitud: Es una isometría seguida de una escalado isotrópico ( un escalado uniforme en toda la imagen ), en este caso las distancias no se mantienen pero si sus proporciones. Los ángulos tampoco son alterados.

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} \lambda\cos\theta & -\lambda\sin\theta & t_x \\ \lambda\sin\theta & \lambda\cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

Afinidad: Las afinidades pueden representarse tal que:

## 2. Detección de características.

	Harris+CN	Harris+CNME	SURF
Intensidad	Si	Si	Si
Isometría	Si	Si	Si
Similitud	No	Pseudo	Si
Afinidad	No	No	Si

Cuadro 2.1.: Resumen de la capacidad de descriptores.

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

La afinidad puede descomponerse tal que :

$$A = UDV^T = (UV^T)(VDV^T) = R(\theta)R(-\alpha)DR(\alpha)$$

, es decir, se compone de rotaciones y de escalados no isotrópicos donde  $D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$ .

La transformación afín entonces es la concatenación de una rotación por  $\alpha$ , un escalado para cada dirección (notar que la figura esta rotada), revertir la rotación de la figura y una nueva rotación por  $\theta$ . Esta transformación conserva el paralelismo de las direcciones, no así los ángulos, se conservan ratios de distancias de líneas paralelas y ratio de áreas.

La detección de Harris se va a mantener respecto de rotaciones y traslaciones, unas transformaciones que impliquen más grados de libertad llevara consigo que el punto interesante se difumine disminuyendo el valor del operador de Harris. Con la aplicación de múltiples escalas se consigue cierta flexibilidad en los cambios sobre la escala. SURF, en cambio, puede resistir afinidades debido a que considera múltiples escalas de diferentes filtros a sobre la imagen pudiendo adaptarse a transformaciones afines, ver cuadro 2.1 .



## 3. Geometría

Una imagen puede considerarse como una medición de luz. En ese proceso se producen diferentes etapas que influyen en el resultado final, en la imagen percibida por la cámara. Este capítulo se centra en la geometría que describe como los rayos de luz forman una imagen, y como un par de imágenes están relacionadas entre si.

Una cámara es un dispositivo que construye imágenes planas del mundo tridimensional a través de una proyección. La geometría de la formación de la imagen fue entendida en el Renacimiento por pintores que utilizaban los puntos en el infinito y otras construcciones geométricas. Con el tiempo surgió la fotografía, se estudio como hacer mediciones en las fotografías mostrando escenas en perspectiva, y esto llevo a la fotogrametría, ciencia con un amplio abanico de aplicaciones. Durante el pasado siglo, los matemáticos han desarrollado la geometría proyectiva que trata sobre el manejo de puntos en el infinito y de proyecciones en perspectiva. Este es el campo científico que sustenta el uso de imágenes en la percepción del entorno.

En este estudio se manejan imágenes obtenidas con cámaras no calibradas. A pesar de eso, es necesario ver los siguientes conceptos para describir cuales son las dificultades que presenta un sistema estéreo de cámaras y valorar las ventajas que tiene la opción elegida en el tratamiento de las imágenes de este sistema.

### 3.1. Modelo de cámara

#### 3.1.1. Geometría Proyectiva y Coordenadas Homogéneas

En la geometría proyectiva no son validas las nociones de distancia o ángulo, es una geometría basada en el estudio de las configuraciones de puntos y líneas. La estructura de las incidencias y el cross-ratio son dos propiedades invariables sobre las transformaciones proyectivas fundamentales para

### 3. Geometría

esta rama. En ellas se encuentra el estudio de los puntos en el infinito cuyo tratamiento es facilitado por las coordenadas homogéneas.

Un punto en un plano puede ser representado por un par de coordenadas  $(x, y)$  en  $\mathbb{R}^2$ .

Representación homogénea de líneas -> Una línea en el plano es representada por una ecuación de la forma  $ax + by + c = 0$ , diferentes valores de  $a$ ,  $b$  y  $c$  generan las diferentes líneas. Entonces se puede representar una línea por el vector  $(a, b, c)$ , sin embargo no hay una correspondencia uno-a-uno entre esta representación y las líneas ya que  $ax + by + c = 0$  y  $(ka)x + (kb)y + (kc) = 0$  son la misma línea. Entonces los vectores  $(a, b, c)$  y  $k(a, b, c)$  representan la misma línea para cualquier  $k$  distinto de cero con lo que representan un grupo de vectores equivalentes, siendo cada vector un vector homogéneo. Cualquier vector  $(a, b, c)$  concreto es representativo de su grupo de equivalencia, y el conjunto de grupos de vectores en  $\mathbb{R}^3 \setminus (0, 0, 0)$  forman el espacio proyectivo  $\mathbb{P}^2$ , ( el vector  $(0, 0, 0)$  no representa ninguna línea y por tanto no está incluido ).

Representación homogénea de puntos -> Un punto  $q = (x, y)$  pertenece a una línea  $l = (a, b, c)$  si y solo si  $ax + by + c = 0$ , esto puede ser escrito en forma de producto interior de dos vectores como  $(x, y, 1) \cdot (a, b, c) = (x, y, 1) \cdot l = 0$ . El punto  $q$  en  $\mathbb{R}^2$  es representado como un vector de 3 componentes, añadiendo una tercera componente igual a 1. Notar que para cualquier  $k$  distinto de cero se mantiene  $(kx, ky, k1) \cdot l = 0 \Rightarrow (x, y, 1) \cdot l = 0$ , entonces podemos considerar el conjunto de vectores  $(kx, ky, k1)$  como representación del punto  $q = (x, y)$ . De esta forma un vector homogéneo cualquiera con la forma  $X = (x_1, x_2, x_3)$  es representación del punto  $(\frac{x_1}{x_3}, \frac{x_2}{x_3})$  en  $\mathbb{R}^2$ , con lo que los puntos pueden ser manejados como vectores homogéneos en  $\mathbb{P}^2$ .

La línea que une dos puntos  $p$  y  $q$  en  $\mathbb{P}^2$  está dada por el producto exterior  $p \times q$ .

#### 3.1.2. Cámara Estenopeica<sup>1</sup>

El modelo matemático de la cámara estenopeica describe la relación entre las coordenadas de puntos en el mundo real descrito en 3 dimensiones, con sus proyecciones en una imagen. Este modelo no tiene en cuenta distorsiones ni deformaciones producidas por lentes, luego solo puede considerarse como una primera aproximación

En el modelo de cámara estenopeica la luz proveniente de la escena pasa por un agujero, el estenopo,

---

<sup>1</sup>La cámara estenopeica es posible que se conozca por sus denominación en inglés "pinhole camera".

### 3. Geometría

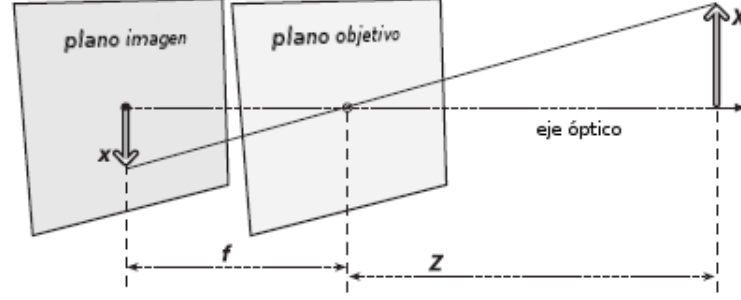


Figura 3.1.1.: Modelo cámara estenopeica

que solo deja pasar un rayo de luz para cada punto de la escena, estos rayos son proyectados en el plano imagen donde se forma la imagen en sí. De esta forma la imagen esta siempre enfocada y el tamaño de la imagen en relación a la escena esta dado por un único parámetro, la longitud focal. En la figura 3.1.1 podemos ver la longitud focal  $f$  que será igual a la distancia desde el estenopo hasta el plano imagen,  $Z$  es la distancia desde el estenopo hasta el objeto,  $X$  es la longitud del objeto y  $x$  es la longitud de la imagen del objeto. De esta forma tenemos que:

$$\frac{-x}{f} = \frac{X}{Z} \iff -x = f \frac{X}{Z}$$

Para facilitar el modelo matemático se considera un sistema “virtual” equivalente al anterior, se coloca el plano imagen entre la escena y el estenopo que pasa a ser el centro de proyección. En este sistema la relación  $x / f = X / Z$  se mantiene librándonos además del signo negativo. En el plano imagen tenemos el punto principal, lo ideal es que este punto sea el centro de la imagen coincidiendo por donde el eje óptico la corta, pero en la realidad esto no es así lo que hace necesario introducir 2 nuevos parámetros,  $c_x$  y  $c_y$  con lo que se registra el posible desplazamiento con respecto al eje óptico. Entonces se modela como un punto  $Q$  en el mundo físico con coordenadas  $(X, Y, Z)$  es proyectado sobre la imagen con coordenadas  $(x_{img}, y_{img})$ , respecto al sistema coordenadas de la imagen:

$$x_{img} = f_x \left( \frac{X}{Z} \right) + c_x, \quad y_{img} = f_y \left( \frac{Y}{Z} \right) + c_y$$

Hay dos longitudes focales,  $f_x$  y  $f_y$ , ya que en las cámaras reales el sensor CCD, que en el modelo es el plano imagen, contienen píxeles rectangulares y no cuadrados, esto supone una diferencias en las proporciones de la imagen que se compensa con estos factores.

Haciendo uso de las coordenadas homogéneas y agrupando nuestros parámetros en una matriz 3x3 se tiene que esta relación puede ser escrita de la siguiente forma:

### 3. Geometría

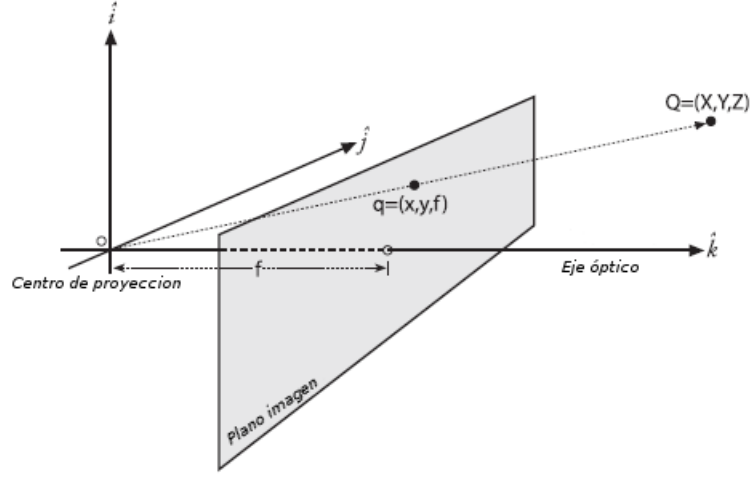


Figura 3.1.2.: Modelo Proyectivo de cámara estepeoneica

$$q = MQ, \quad \text{donde} \quad q = \begin{bmatrix} x \\ y \\ w \end{bmatrix}, \quad M = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, \quad Q = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

En estas ecuaciones  $w = Z$ , siendo  $w$  la componente *proyectiva* de las coordenadas homogéneas del punto  $q$ .

#### 3.1.3. Distorsión de las lentes

Al modelo anterior hay que añadirle más parámetros ya que en la obtención de imágenes por cualquier cámara actual se hace uso de lentes. Las propias características de las lentes y la imperfección en la colocación en la mecánica de la cámara producen nuevas modificaciones en la trayectoria de los rayos de luz.

Por un lado esta la distorsión radial provocada por la diferencia en la inclinación con la que los rayos de luz llegan al objetivo de la cámara, esto hace que la cantidad de rayos de luz que generan diferentes partes de la imagen sea diferente, en el centro serán recibidos más rayos de luz que cerca de los bordes.

La distorsión tangencial esta producida por la inclinación del plano que finalmente es sensor de la luz, el sensor CMOS donde se encuentran las celdas fotosensibles, con el plano de la lente. Esta distorsión esta muy relacionado con la precisión en el proceso de fabricación de la cámara.

Este tipo de distorsiones pueden ser modeladas de diferentes formas, en las librerías OpenCV se sigue el modelo descrito en [10], donde teniendo un punto en la imagen  $p = (x, y, 1)$  sera modificado

### 3. Geometría

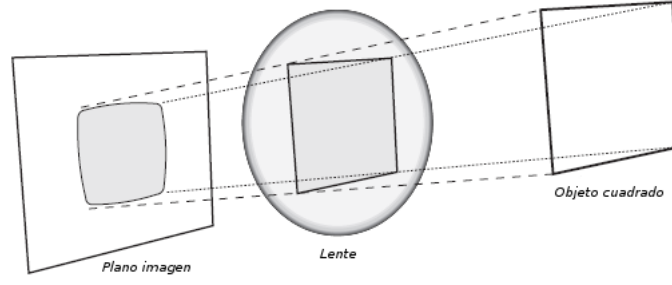


Figura 3.1.3.: Distorsión radial generada por la distinta inclinación con la que los rayos atraviesan la lente en diferentes zonas.

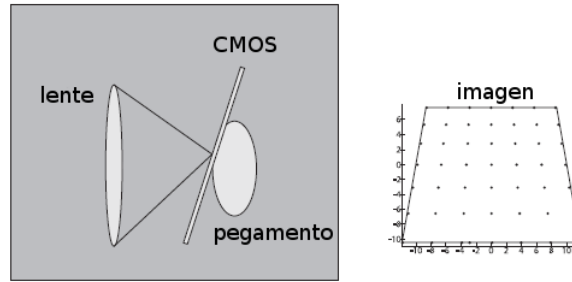


Figura 3.1.4.: Distorsión tangencial producida por la inclinación del sensor respecto de la lente.

por la distorsión a  $p' = (x', y', 1)$  tal que :

$$x' = x(1 + k_1r^2 + k_2r^4 + k_3r^6) + 2h_1x'y' + h_2(r^2 + 2x^2)$$

$$y' = y(1 + k_1r^2 + k_2r^4 + k_3r^6) + 2h_2x'y' + h_1(r^2 + 2y^2)$$

Donde :

$$r^2 = x^2 + y^2$$

Los parámetros  $k_1, k_2, k_3$  establecen la distorsión radial,  $h_1, h_2$  la tangencial. En las imágenes utilizadas en este estudio no trataremos estas distorsiones explícitamente, aunque hay que tener en cuenta que si estas distorsiones son muy acentuadas en las imágenes objeto de rectificación, la corrección de estas distorsiones será un paso ineludible y previo al cálculo de la geometría epipolar.

#### 3.1.4. Calibración

Todos estos parámetros modelizan los cambios que los rayos de luz pueden sufrir en el proceso de obtención de una imagen con respecto al modelo ideal de cámara estenopeica. Cada cámara tendrá

### 3. Geometría

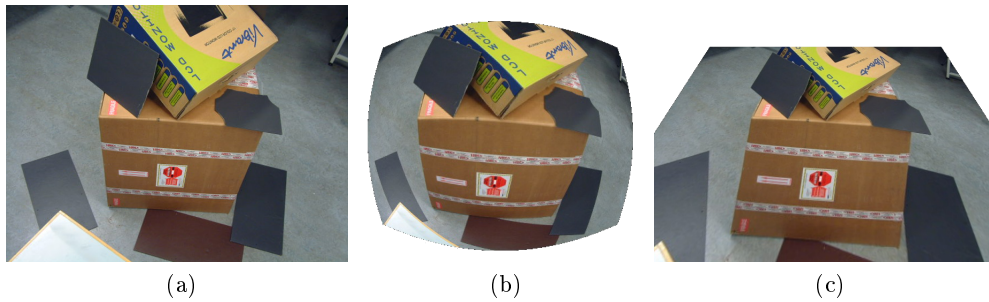


Figura 3.1.5.: Efectos de las distorsión radial y tangencial en una imagen real modificada artificialmente.

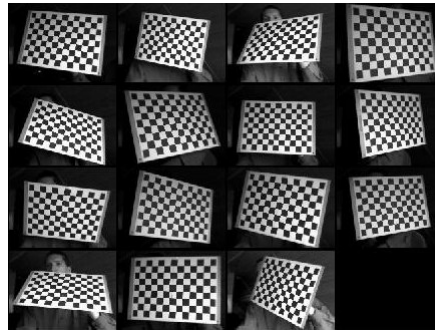


Figura 3.1.6.: Calibración una cámara capturando un objeto conocido en diferentes posiciones.

parámetros diferentes, por esto es necesario la “calibración” de las cámaras si se quiere conseguir más confianza en las mediciones realizadas sobre las imágenes.

En el proceso de calibración se obtienen imágenes de un objeto en diferentes posiciones respecto de la cámara, las dimensiones del objeto se conocen (lo más común es un tablero con el patrón del juego de ajedrez), con un numero suficiente de imágenes pueden deducirse los parámetros intrínsecos y de distorsión de la cámara. Puede llegar a ser tedioso calibrar las cámaras en cada uso, además este proceso no esta exento de error ya que se alcanza un compromiso entre los diferentes parámetros por métodos numéricos. El más utilizado es el método implementado en OpenCV, descrito en la publicación [9], en el que calcula las longitudes focales y los desplazamientos del punto principal de la imagen respecto del eje óptico, y para las distorsiones utiliza el método descrito en [10].

No entraremos en más detalles dado que se van a realizar rectificación de imágenes no calibradas evitando la necesidad de realizar este proceso.

## 3.2. Introducción a la Geometría Epipolar

La geometría básica de un sistema de cámaras estéreo es referida como geometría epipolar. En esencia esta geometría combina dos cámaras estenopeicas, una para cada cámara, y los puntos llamados “epipolos”. Para explicar su utilidad, primero se define alguna terminología nueva, entonces se puede ver fácilmente que esa geometría limita considerablemente las regiones de la imagen en la que se pueden encontrar equivalencias entre los puntos de las 2 imágenes.

Ilustrado en la figura 3.2.1, el punto  $P$  es un punto cualquiera, los puntos  $O$ ,  $O'$  son los centros focales de cada cámara, estos 3 puntos forman el plano epipolar al que también pertenecen las imágenes del punto  $P$ , que son  $p$  y  $p'$ . En concreto el punto  $p$  se encuentra en la línea generada uniendo  $P$  y  $O$ , de manera similar el punto  $p'$  se encuentra en la línea generada uniendo  $P$  y  $O'$ . El plano epipolar corta a los planos imagen  $\Pi$  y  $\Pi'$  en las líneas  $l$  y  $l'$ , además la línea que une los centros focales  $O$  y  $O'$  la llamaremos línea de base, esta corta a los planos imagen  $\Pi$  y  $\Pi'$  en los puntos  $e$  y  $e'$  denominados epipolos. La línea  $l$  es la línea epipolar asociada al punto  $p$  que pasa por el punto  $e$ , y la línea  $l'$  es la línea epipolar asociada al punto  $p'$  que pasa por el punto  $e'$ .

Esta geometría establece restricciones entre las diferentes entidades, punto y línea, son las siguientes:

- Cada punto tridimensional visto por el par de cámaras pertenece al plano epipolar que corta a cada imagen en una línea epipolar.
- Dado un punto en una imagen, su punto equivalente en la otra imagen debe pertenecer a la correspondiente línea epipolar, esto es conocido como la restricción epipolar.
- La restricción epipolar supone reducir la búsqueda del punto equivalente en la otra imagen de una búsqueda bidimensional a unidimensional a lo largo de la línea epipolar. Esto elimina esfuerzo de computación y permite desechar gran cantidad de puntos que de otra manera generarían correspondencias erróneas.
- El orden se mantiene. Si los puntos A, B y C son vistos por ambas cámaras, y están dispuestos horizontalmente respecto de una de ellas, en la otra también ocurrirán horizontalmente. Puede que debido a oclusiones presentes en una cámara y en la otra no, el conjunto de puntos no sea el mismo.

### 3. Geometría

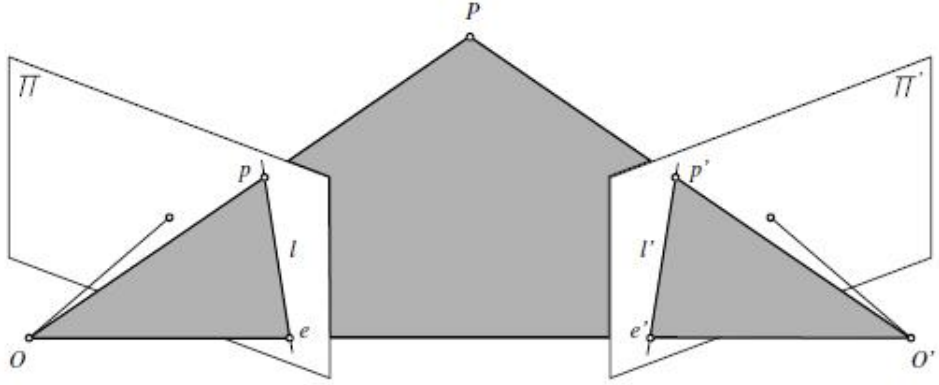


Figura 3.2.1.: Sistema de cámaras estéreo y su geometría epipolar.

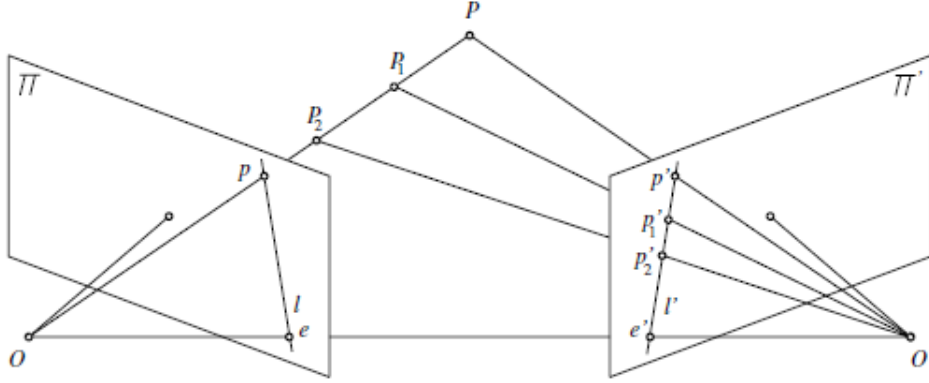


Figura 3.2.2.: Posibles hipótesis para un punto basadas en la línea epipolar .

Para ilustrar estos hechos se toma el ejemplo de la figura 3.2.2, en ella se puede apreciar como la línea epipolar  $l'$  es la imagen del rayo que va desde  $P$  hasta  $O$ , y por lo tanto  $p$  puede ser la imagen de cualquier punto de este rayo, por lo que el punto equivalente tiene que estar contenido en la línea epipolar  $l'$ .

La geometría epipolar se expresa algebraicamente según dos casos diferentes, si tenemos las cámaras calibradas o no. En el caso de que tengamos las cámaras calibradas se habla de matriz esencial, en caso contrario de matriz fundamental.



### 3. Geometría

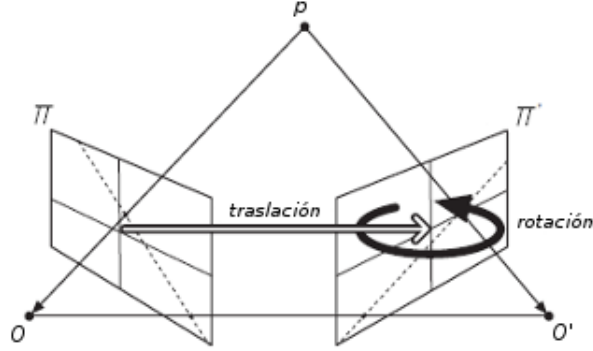


Figura 3.2.3.: La matriz esencial contiene información sobre la rotación y la traslación entre las cámaras.

#### 3.2.1. Matriz esencial

La matriz esencial aparece cuando se tienen imágenes calibradas, en este caso existe un sistema de referencia común y puede decirse que los puntos están expresados en un sistema homogéneo de coordenadas normalizadas. La matriz que relaciona los puntos entre las 2 imágenes no contiene ninguna información acerca de los parámetros intrínsecos de la proyección y es llamada matriz esencial. Esta matriz expresa la transformación necesaria para alinear las imágenes por medio de una rotación y una traslación describiendo así la relación entre las posiciones de las cámaras en el mundo físico.

Un sistema de coordenadas normalizadas expresa que teniendo calibrada una cámara con su matriz de proyección  $P = M[R|T]$  y varios puntos  $q = PQ$ , si la matriz de parámetros intrínsecos  $M$  es conocida, entonces se puede aplicar su inversa tal que  $\bar{q} = M^{-1}q$  donde  $\bar{q}$  es un punto de la imagen en coordenadas normalizadas, siendo su proyección  $\bar{q} = [R|T]Q$ . De esta forma se elimina el efecto de la matriz de calibración. Luego se define la matriz esencial en términos de las coordenadas normalizadas imagen con las correspondencias entre puntos de dos imágenes  $q \rightarrow q'$  tal que :

$$\bar{q}'^T E \bar{q} = 0$$

Que si sustituimos  $\bar{q}$  y  $\bar{q}'$  se tiene :

$$q'^T M'^{-T} E M^{-1} q = 0$$

A partir de la cual se relacionan fácilmente la matriz esencial y la fundamental ya que la matriz fundamental cumple que  $q'^T F q = 0$ , por tanto  $E = M'^T F M$ .

### 3. Geometría

#### 3.2.2. Calibración de un Sistema Estéreo

La calibración de un sistema estéreo consiste en conocer la rotación y la traslación que relacionan el par de cámaras de nuestro sistema, se sigue el mismo procedimiento en el que se capta la imagen de un objeto con un patrón conocido en diferentes posiciones. A partir de las imágenes captadas por el par de cámaras hay que calcular la geometría epipolar que en este caso será en forma de matriz esencial.

Otro método es el denominado auto-calibración que determina los parámetros del sistema directamente de correspondencias de puntos entre múltiples imágenes no calibradas calculando la matriz fundamental sobre diferentes posiciones del sistema en la escena. Este método tiene la ventaja de poder calibrar el sistema sin necesitar un patrón de calibración, pudiendo calibrarse a partir de la propia secuencia de imágenes percibida. Esta funcionalidad resulta muy interesante en sistemas autónomos permitiendo mayor flexibilidad en los tiempos de operación y ampliando las posibilidades de acción. Como desventaja esta que se desconoce el factor de escala, es decir, no se conoce cual es la relación de tamaño entre la escena y el plano imagen. Además es interesante expresar el modelo sin relación a las coordenadas “externas” que pueden variar por diversos motivos.

#### 3.2.3. Matriz fundamental.

La matriz fundamental relaciona los puntos de una imagen expresados en el sistema de coordenadas de píxeles de una imagen en el sistema de coordenadas de píxeles de la otra imagen, conteniendo así cierta información sobre los parámetros intrínsecos y de distorsión de cada cámara.

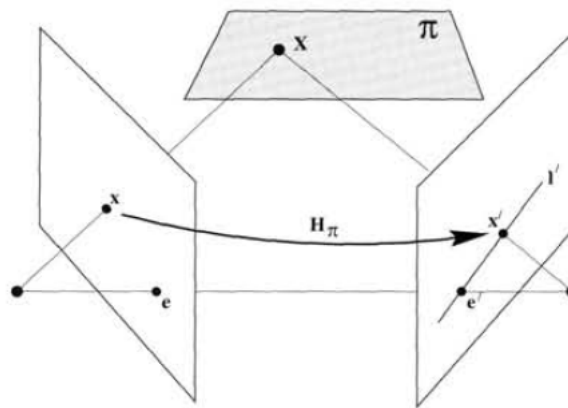


Figura 3.2.4.: La transferencia de puntos entre las dos imágenes a través de un plano es la matriz fundamental.

### 3. Geometría

Dadas dos imágenes  $I$  e  $I'$  de una escena, tenemos dos conjuntos de puntos  $x$  y  $x'$  que son las proyecciones de un conjunto de puntos  $X$  se puede construir la matriz fundamental. Esta matriz es una homografía tal que :

$$x' = H_{\Pi} x$$

Las líneas de  $I'$  se pueden describir como<sup>2</sup>:

$$l' = e' \times x' = [e']_{\times} x'$$

Sustituyendo la primera en la segunda:

$$l' = [e']_{\times} H_{\Pi} x = F x$$

Donde se define la matriz fundamental como  $F = [e']_{\times} H_{\Pi}$ . Como  $[e']_{\times}$  tiene rango 2 y  $H_{\Pi}$  tiene rango 3,  $F$  será una matriz de rango 2 que además debe satisfacer la condición de correspondencia:

$$x'^T F x = 0$$

Dadas al menos 8 correspondencias entre puntos de  $x$  y  $x'$  la matriz  $F$  puede ser calculada. Como se ha dicho anteriormente la matriz fundamental relaciona puntos en una imagen con líneas en la otra, cumpliéndose además que:

$$F e = 0 = F^T e'$$

Donde  $e = (e_u, e_v, 1)^T$  es la imagen en  $I$  del centro de proyección  $O'$ , y  $e' = (e'_u, e'_v, 1)^T$  es la imagen en  $I'$  del centro de proyección  $O$ . Como la matriz fundamental es de rango 2, esto hace que de los 9 componentes de la matriz tenga 7 grados de libertad, 2 para cada epipolo y 3 para la homografía que relaciona los dos planos imagen, el factor de escala no se tiene en cuenta.

---

<sup>2</sup>La forma matricial del producto exterior de un vector  $v = (v_1, v_2, v_3)$  sera la matriz antisimetrica

$$[v]_{\times} = \begin{bmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{bmatrix}$$

### 3. Geometría

#### Computación de la Matriz Fundamental

La computación de la matriz fundamental requiere de un cuidadoso manejo de las operaciones ya que la precisión cuando se manejan valores que representan el infinito, como es el caso de puntos en coordenadas homogéneas, puede verse afectada de manera significativa. En el cálculo de esta matriz se utiliza el método normalizado de 8-puntos que con el que se obtienen los mejores resultados según las publicaciones [12, 14] y en las que se pueden ver todos los detalles del algoritmo.

El planteamiento es : dadas  $n \geq 8$  correspondencias entre puntos de 2 imágenes  $m_i \leftrightarrow m'_i$  calcular la matriz  $F$  tal que  $m'^T F m = 0$  .

1. Normalización: Transformar las coordenadas de las imágenes con una traslación y un escalado de tal forma que el centroide de los puntos este en el origen de las coordenadas y que la distancia media cuadrática<sup>3</sup> de los puntos al origen sea igual a  $\sqrt{2}$ , es decir, se aplican unas transformaciones tal que :  $\hat{m} = Tm$  y  $\hat{m}' = T'm'$ .
2. Calcular la solución lineal:

$$\hat{m}'^T \hat{F} \hat{m} = 0$$

Donde  $\hat{m} = (x, y, 1), \hat{m}' = (x', y', 1), \hat{F} = \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}$  entonces para cada correspondencia se tiene :

$$x'xf_{13} + x'yf_{12} + x'f_{13} + y'xf_{23} + y'yf_{22} + y'f_{23} + xf_{31} + yf_{32} + f_{33} = 0$$

Crear un sistema  $Af = 0$  de la forma:

---

<sup>3</sup>Media cuadrática ( del ingles Root Mean Square ) :

$$X_{RMS} = \sqrt{\frac{1}{N} \sum_{i=0}^N x_i^2} = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_N^2}{N}}$$

### 3. Geometría

$$\begin{bmatrix} x'_1x_1 & x'_1y_1 & x'_1 & y'_1x_1 & y'_1y_1 & y'_1 & x_1 & y_1 & 1 \\ x'_2x_2 & x'_2y_2 & x'_2 & y'_2x_2 & y'_2y_2 & y'_2 & x_2 & y_2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x'_nx_n & x'_ny_n & x'_n & y'_nx_n & y'_ny_n & y'_n & x_n & y_n & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{bmatrix} = 0$$

La solución a este sistema de ecuaciones homogéneo es resuelto en función de una escala, para que exista solución la matriz  $A$  debe tener al menos rango 8 . Las correspondencias suelen estar afectadas por errores en la medición ( ruido ) lo que provoca que pueda tener rango 9, en este caso se encuentra la solución por mínimos cuadrados que minimiza  $\|Af\|$  condicionado a que  $\|f\| = 1$  mediante la descomposición de valores singulares. Esta solución para  $f$  será el vector singular correspondiente al menor valor singular de  $A$ , es decir, la ultima columna de  $V$  en la descomposición  $A = UDV^T$ .

3. Forzar a la matriz fundamental la restricción de rango 2 ; Esta es una propiedad importante de la matriz fundamental, recordar que el núcleo izquierdo y derecho representan los epipolos de las dos imágenes, si la matriz fundamental no tuviera rango 2 las matrices epipolares no coincidirían. Para conseguir esto “corregir” la matriz  $\hat{F}$  encontrada en el apartado anterior por  $\hat{F}'$  obtenida utilizando de nuevo la descomposición de valores singulares. Teniendo  $\hat{F} = UDV^T$ , donde  $D$  es una matriz diagonal  $D = \text{diag}(r, s, t)$  que cumple  $r \geq s \geq t$  entonces establecer  $\hat{F}' = U \text{diag}(r, s, 0)V^T$ , de esta forma  $F'$  minimiza  $\|F - F'\|$  consiguiendo así la matriz más cercana en términos de la norma matricial.
4. Por ultimo revertir la normalización del sistema de coordenadas en el que se ha calculado la matriz fundamental, para eso hay que aplicar  $F = T'^T \hat{F}' T$ .

La normalización de los puntos no es opcional, para una justificación detallada mirar [14], solo decir que los resultados son mucho más precisos con la normalización debido en esencia a la diferencia

### 3. Geometría

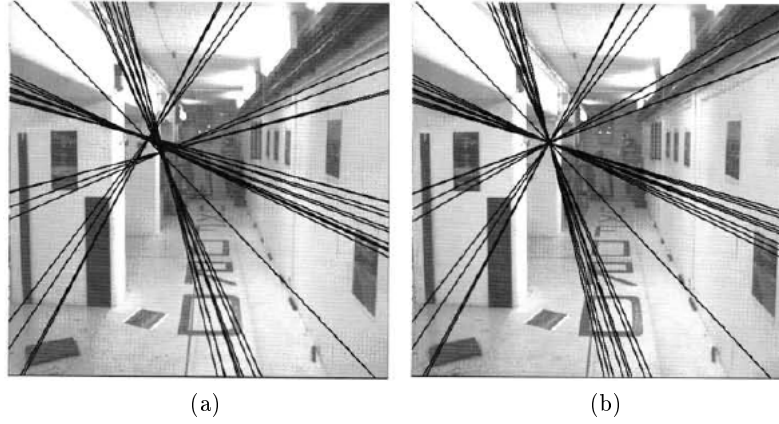


Figura 3.2.5.: Imprecisión de la matriz fundamental.

de magnitudes en las coordenadas de un punto. Sin normalización un punto cualquiera podría ser  $(x, y, w)^T = (100, 123, 1)^T$ , donde las coordenadas  $x$  e  $y$  son mucho más grandes que  $w$ , entonces las entradas de la matriz  $A$  del sistema de ecuaciones afectarían de diferente forma a norma matricial. Este efecto se magnifica más en presencia de ruido en los datos y es por esto que la normalización genera resultados mucho más precisos.

En las figura 3.2.5 se pueden ver líneas epipolares no coincidentes por el efecto del rango 3 en la matriz singular, no encuentran epipolo común, en la imagen de la izquierda (a). A la derecha (b) se aprecia el efecto de forzar el rango 2 por el método descrito utilizando la descomposición de valores singulares. La presencia del epipolo sobre la imagen puede producirse en dos imágenes consecutivas con una traslación hacia delante, o en sistema estéreo con un ángulo cerrado entre ellas.

### 3.3. Rectificación de imágenes

En los sistemas estéreo de cámaras es difícil conseguir una configuración de cámaras perfectamente alineadas, eso significa que físicamente los planos imagen de cada cámara son perfectamente co-planares y además con las filas alineadas. Debido a que en la práctica es imposible conseguir esto, se opta por alterar las imágenes para que aparezcan como si hubieran sido tomadas por dos cámaras perfectamente alineadas horizontalmente. Con esta rectificación los ejes ópticos de ambas cámaras son paralelos y se considera que se encuentran en el infinito.

Para conseguir esta configuración entre las imágenes se sigue un proceso diferente ante un sistema

### 3. Geometría

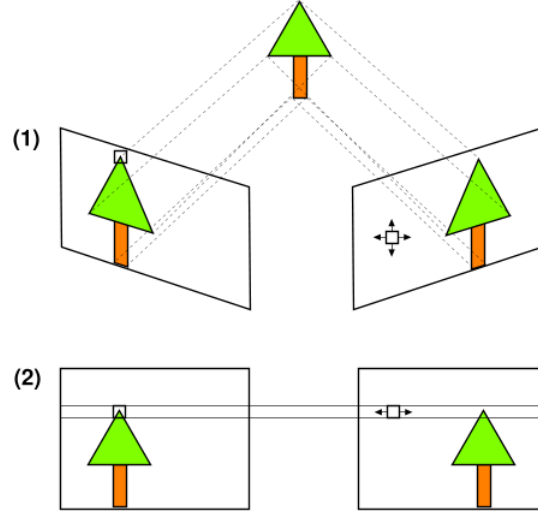


Figura 3.3.1.: El objetivo de la rectificación es alinear las filas de las 2 imágenes.

calibrado, en el que las coordenadas son comunes entre las imágenes y se maneja una matriz esencial, o si se trata de imágenes no calibradas, en el que cada imagen tiene su propio sistema de coordenadas expresando su relación con una matriz fundamental.

En un sistema calibrado la matriz esencial se reduce a  $sRT$ , una factor de escala, una rotación y una traslación, la estrategia de rectificación puede ser entonces repartir la rotación entre las cámaras, como en el algoritmo de Bouguet [16].

Al rectificar un par de imágenes no calibradas, y por tanto ausencia de coordenadas comunes entre las imágenes, se añaden grados de libertad a las rectificaciones posibles, este es caso estudiado.

En la rectificación no calibrada se colocan los epipolos en el infinito, consiguiendo que las líneas epipolares sean paralelas entre ellas, además se consigue que los puntos equivalentes tengan la misma coordenada  $y$  al llevar los epipolos en la dirección  $(1, 0, 0) \simeq (e_u, 0, 0)$ . La matriz fundamental ideal entre dos imágenes que cumplan lo anterior sería :

$$\bar{F} = [I]_{\times} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$$

Donde  $[I]_{\times}$  significa la forma matricial del producto exterior del vector  $I = (1, 0, 0)$ . Las homografías de rectificación dan las nuevas coordenadas como  $\bar{m} = Hm$  y  $\bar{m}' = H'm'$ . En este caso se tiene que  $\bar{m}'^T \bar{F} \bar{m} = 0$  y que  $m'^T H'^T \bar{F} H m = 0$  luego tenemos que  $H'^T \bar{F} H = F$ .

### 3. Geometría

El par de homografías que satisfacen esta ultima ecuación no son únicas, y es por eso que hay diferentes aproximaciones intentando reducir las distorsiones que se introducen en este proceso..

#### 3.3.1. Rectificación de Hartley

Para alinear las filas mediante transformaciones proyectivas entre dos imágenes a partir de la matriz fundamental, Hartley [17], lo primero que hace es calcular los epipolos a partir de la matriz fundamental, esto es fácil ya que  $Fe = 0$  ,  $(e')^T F = 0$  . Entonces se establece la homografía  $H'$  de tal forma que lleve el epipolo derecho al punto 2D homogéneo  $(1, 0, 0)^T$ , suponiendo que el epipolo pertenece al eje X tal que  $(e')^T = (f, 0, 1)$  , la matriz que envía el epipolo al infinito seria :

$$G = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1/f & 0 & 1 \end{pmatrix}$$

Se puede comprobar fácilmente que la matriz  $G$  envía el punto  $e'$  al infinito.

Ya que la homografía tiene 7 grados de libertad ( sin contar la escala ), y 3 son utilizados para relacionar las coordenadas con el infinito, quedan 4 grados de libertad para elegir esta primera homografía, estos 4 grados son prácticamente libres luego hay que hacer una buena elección en la que se minimicen las distorsiones. Para encontrar una buena  $H'$  se elige un punto en la imagen donde minimizar especialmente la distorsión, la idea es que la homografía se comporte solo como rotación y traslación alrededor de ese punto todo lo que se pueda, una elección razonable seria el centro de la imagen. Después, para el punto elegido en la imagen derecha hay que calcular la traslación  $T$  que lo lleve hasta el origen y la  $R$  que lo lleve al epipolo  $(e')^T = (e'_u, 0, 1)$  . Entonces la homografía resultante sera  $H' = GRT$  donde :

$$d = \sqrt{(e'_u)^2 + (e'_v)^2}$$

$$\alpha = \frac{e'_u}{d} \quad ; \quad \beta = \frac{e'_v}{d}$$



### 3. Geometría

$$R = \begin{pmatrix} \alpha & \beta & 0 \\ -\beta & \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$T = \begin{pmatrix} 1 & 0 & -c_x \\ 0 & 1 & -c_y \\ 0 & 0 & 1 \end{pmatrix}$$

Luego, para poder visualizar la imagen rectificadora hay que recolocar las coordenadas de la imagen revirtiendo el efecto de  $T$  aunque esto no afectara a las propiedades de la imagen resultante.

#### Homografía Pareja

Con homografía de rectificación  $H'$ , ahora queda encontrar la homografía  $H$  que sea compatible con la anterior, es decir que cumpla  $m'^T H'^T F H m = 0$ . En el apéndice A.4 esta desarrollada la deducción de la expresión de la homografía buscada que será:

$$H = (I - H'e'a^T)H'M$$

para algún vector  $a$ . Entonces  $H$  tiene colocara el epipolo izquierdo en el infinito y hará que las filas de las dos imágenes estén alineadas utilizando 3 grados de libertad como la homografía derecha.

#### Minimización de Distorsiones

Al llevar el epipolo  $e'$  al punto en el infinito  $(1, 0, 0)^T$ , supone que  $(I + H'e'a^T) = I + (1, 0, 0)^T a^T$  que se puede escribir en la forma :

$$A = \begin{bmatrix} a & b & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

### 3. Geometría

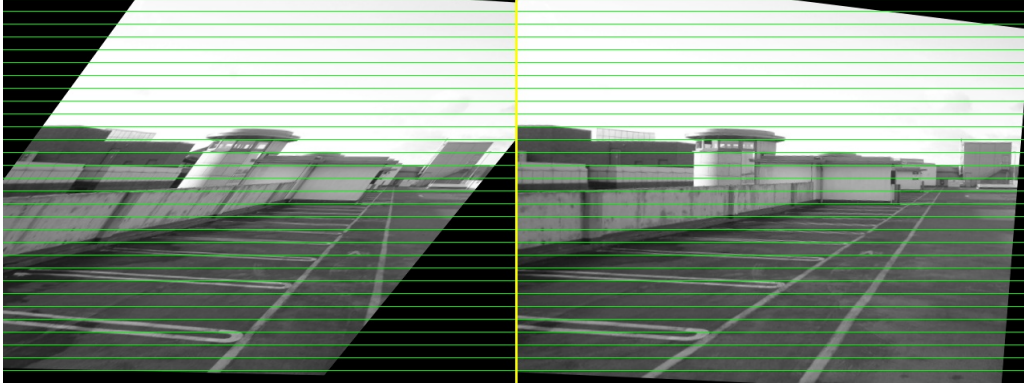


Figura 3.3.2.: Torsión introducida en la imagen izquierda en la rectificación.

Esto se produce porque la matriz fundamental tiene rango 2, es decir, no contiene información sobre la posición de la coordenada  $x$ , lo que hace que la primera fila de nuestras homografías de rectificación queden a nuestra elección, por esto podemos establecer la primera fila en cada homografía sin invalidar las ecuaciones que hemos utilizado para calcularlas.

$$AHe = He$$

Y para las equivalencias ya rectificadas:

$$H'^T A'^T \bar{F} A H = K'^T \bar{F} K = F \quad \text{donde} \quad K = A H$$

Para elegir la matriz  $A$  Hartley considera la minimización de la distancia total entre todas las equivalencias entre las imágenes. Es decir, encontrar una  $H$  que minimiza  $\sum_n d(AH p_n^i, H' p_n^d)^2$ , esto es un sistema de ecuaciones lineal que puede resolverse por mínimos cuadrados, estas dos homografías determinan la rectificación estéreo.

Como ya se ha comentado, al aplicar las homografías de rectificación se introducen distorsiones que provocan distorsiones en las imágenes, pero la interpretación de “distorsión” cambia entre diferentes autores. En particular la función que Hartley minimiza puede provocar “torsiones” en las imágenes fácilmente apreciables, ver figura 3.3.2 .

Por esa razón hay otras consideraciones a tener en cuenta en la elección de la matriz  $A$ , Zhang [19] intenta mantener la apariencia de la imagen. Considerando la distorsión como una modificación de los ángulos entre las líneas y cambio en la relación de aspecto de las imágenes, esto es, intenta mantener las cualidades afines de las imágenes..

### 3. Geometría

La elección de la matriz de torsión se hace para cada imagen independientemente, por ahora dejamos  $c = 0$  ya que es un componente de traslación que no influye en esta fase. La distorsión de la rectificación se puede medir sobre las líneas que parten en dos a cada imagen, con los puntos  $p_1 = (w/2, 0, 1)^T$ ,  $p_2 = (w, h/2, 1)^T$ ,  $p_3 = (w/2, h, 1)^T$ ,  $p_4 = (0, h/2, 1)^T$ , donde  $w$  es el ancho de la imagen y  $h$  es la altura de la imagen. Estos puntos los rectificamos con la homografía  $H$  (con la otra imagen sería  $H'$ ) que tenemos hasta ahora tal que tenemos los puntos  $\hat{p}_1 = Hp_1$ ,  $\hat{p}_2 = Hp_2$ ,  $\hat{p}_3 = Hp_3$ ,  $\hat{p}_4 = Hp_4$ . Entonces se forman los vectores :

$$X = (x_u, y_v)^T = \hat{p}_2 - \hat{p}_4$$

$$Y = (x_u, y_v)^T = \hat{p}_1 - \hat{p}_3$$

Luego para conservar la apariencia de la imagen intentamos preservar la perpendicularidad y la relación de aspecto para estas líneas, consiguiéndolo también para la imagen entera. Estos dos vectores pertenecen al espacio euclídeo por tanto la perpendicularidad se conserva cuando:

$$(AX)^T(AY) = 0$$

La relación de aspecto se conserva si:

$$\frac{(AX)^T(AX)}{(AY)^T(AY)} = \frac{w^2}{h^2}$$

Estas ecuaciones representan polinomios cuadráticos en  $a$  y  $b$  que son las incógnitas buscadas de  $A$ , ecuaciones que se deben satisfacer simultáneamente. Utilizando el método descrito en [11] se obtienen las soluciones :

$$a = \frac{h^2 x_v^2 + w^3 y_v^2}{hw(x_v y_u - x_u y_v)} \quad y \quad b = \frac{h^2 x_u x_v + w^2 y_u y_v}{hw(x_u y_v - x_v y_u)}$$

Finalmente seleccionamos el parámetro de  $c$  que faltaba para el punto seleccionado

#### 3.3.2. Rectificación de Whelan

Paul Whelan [18] presenta un método basado en parte de los anteriores, por un lado elige la primera homografía de rectificación como lo hace Hartley, esta homografía introduce la mínima distorsión cuanto más lejos este la imagen de su epipolo. Por otro lado, utiliza el concepto de distorsión de Zhang, es decir, mantener la perpendicularidad y la relación de aspecto invariables sobre transformaciones afines.

### 3. Geometría

Considera estas cualidades en regiones locales sobre toda la imagen en aproximaciones de primer orden a través de la descomposición de valores singulares. Este método tiene la ventaja en que únicamente esta basado en la matriz fundamental, esto establece una cota superior para el error de rectificación igual al error de la matriz fundamental.

En primer lugar se establece una homografía de manera similar a Hartely :

$$H = \begin{bmatrix} 1 & 0 & 0 \\ -e_v/e_u & 1 & 0 \\ -1/e_u & 0 & 1 \end{bmatrix}$$

Para la *homografía pareja* (recordar que la primera fila no esta determinada) entonces se puede escribir:

$$H' = \begin{bmatrix} 1 & 0 & 0 \\ h'_{21} & h'_{22} & h'_{23} \\ h'_{31} & h'_{32} & h'_{33} \end{bmatrix}$$

De esta forma, desarrollando la ecuación:

$$H'^T \bar{F} H = F$$

Queda :

$$\begin{bmatrix} 1 & h'_{21} & h'_{31} \\ 0 & h'_{22} & h'_{32} \\ 0 & h'_{23} & h'_{33} \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ m & 1 & 0 \\ n & 0 & 1 \end{bmatrix} = \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}$$

$$\begin{bmatrix} (mh'_{31} - nh'_{21}) & h'_{31} & -h'_{21} \\ (mh'_{32} - nh'_{22}) & h'_{32} & -h'_{22} \\ (mh'_{33} - nh'_{23}) & h'_{33} & -h'_{23} \end{bmatrix} = \alpha \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}$$

A partir del cual se puede crear un sistema del tipo  $Ax = 0$  resuelto por mínimos cuadrados por medio de la descomposición por valores singulares.

### 3. Geometría

$$\begin{bmatrix} -n & 0 & 0 & m & 0 & 0 & -f_{11} \\ 0 & 0 & 0 & 1 & 0 & 0 & -f_{12} \\ -1 & 0 & 0 & 0 & 0 & 0 & -f_{13} \\ 0 & -n & 0 & 0 & m & 0 & -f_{21} \\ 0 & 0 & 0 & 0 & 1 & 0 & -f_{22} \\ 0 & -1 & 0 & 0 & 0 & 0 & -f_{23} \\ 0 & 0 & -n & 0 & 0 & m & -f_{31} \\ 0 & 0 & 0 & 0 & 0 & 1 & -f_{32} \\ 0 & 0 & -1 & 0 & 0 & 0 & -f_{33} \end{bmatrix} \begin{bmatrix} h'_{21} \\ h'_{22} \\ h'_{23} \\ h'_{31} \\ h'_{32} \\ h'_{33} \\ \alpha \end{bmatrix} = 0$$

Para la reducción de la distorsión, teníamos:

$$H'^T A'^T \bar{F} A H = K'^T \bar{F} K = F$$

Donde  $A$  y  $A'$  son de la forma:

$$\begin{bmatrix} a & b & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Para minimizar la distorsión se buscan independientemente matrices que minimicen la creación o eliminación de píxeles sobre la imagen. Tras aplicar nuestra rectificación a una imagen estos cambios pueden ser cuantificados en la vecindad de un punto por cualquier norma de la jacobiana:

$$J(K, p) = \begin{bmatrix} \frac{\partial \bar{x}}{\partial x} & \frac{\partial \bar{x}}{\partial y} \\ \frac{\partial \bar{y}}{\partial x} & \frac{\partial \bar{y}}{\partial y} \end{bmatrix}$$

Teniendo como  $\sigma_1(J)$  y  $\sigma_2(J)$  a los valores singulares de la jacobiana en orden decreciente, una transformación ortogonal ideal, que no crea ni destruye píxeles, tendrá sus valores singulares igual a 1, se puede de esta manera cuantificar la creación de píxeles,  $\sigma_1(J) > 1$ , y la eliminación,  $\sigma_1(J) < 1$ .

Lo siguiente es realizar una búsqueda sobre los parámetros  $a$  y  $b$  que mantengan la ortogonalidad y la perspectiva de la imagen original, notar que  $c$  no influye en las propiedades que se intentan preservar por lo que no es necesario en esta fase, posteriormente puede establecerse de tal forma que centre la imagen rectificada en las coordenadas de la imagen original.

### 3. Geometría

El teorema de Wielandt-Hoffman<sup>4</sup>, para valores singulares afirma que si  $A$  y  $E$  son matrices de  $\mathbb{R}^{m \times n}$  con  $m \geq n$ , entonces:

$$\sum_{k=1}^n (\sigma_k(A + E) - \sigma_k(A))^2 \leq \|E\|_F^2$$

Donde  $\|E\|_F$  es la norma de Frobenius de  $E$ . Se establece que si  $A$  es perturbada por  $E$ , la correspondiente perturbación en cualquiera de los valores singulares de  $A$  será menor que el valor de la norma de Frobenius de  $E$ , haciendo posible una búsqueda a través de técnicas iterativas sobre el valor de los valores singulares de la matriz jacobiana sobre varios puntos en la imagen. Cuantificamos estos valores en una malla de puntos distribuida uniformemente sobre la imagen siendo la función final:

$$f(a, b) = \sum_{i=1}^n [(\sigma_1(J(K, p_i)) - 1)^2 + (\sigma_2(J(K, p_i)) - 1)^2]$$

Este funcional se minimiza utilizando el método simplex de Nelder y Mead, mirar A.5, con el que hay que tener cuidado con los diferentes parámetros en su configuración que pueden afectar a los resultados. La matriz jacobiana es aproximada con diferencias centrales finitas sobre cada punto.

---

<sup>4</sup>Matrix Computations, pag. 450, teorema 8.6.4

## 4. Resultados

A continuación se presentan los resultados obtenidos al implementar los métodos explicados a lo largo de esta memoria. En primer lugar los resultados obtenidos con la detección de esquinas con los métodos de SURF y Harris+CN y +CNME, resultados basados en el error de reproyección de la matriz fundamental.

Son utilizadas 6 pares de imágenes estéreo, cada par pertenece a una escena diferente, y para cada par se dispone de correspondencias manualmente seleccionadas que servirán de referencia para calcular el error de las correspondencias calculadas automáticamente en la primera fase, y en la segunda servirán para calcular la matriz fundamental base del proceso de rectificación.

### 4.1. Medición de la calidad de las correspondencias

Para la medición de la calidad de las correspondencias se sigue el siguiente procedimiento:

- Detección de características en cada imagen para cada par.
- Comparación de todas con todas las características detectadas en cada imagen.
- Selección de las 50 correspondencias que alcancen mayor puntuación en su comparación.
- Cálculo de la matriz fundamental por medio de RANSAC y el algoritmo de 8 puntos, en la que se selecciona la matriz fundamental que minimiza el error de reproyección sobre las correspondencias introducidas.
- Cálculo del error de reproyección de la matriz fundamental obtenida en el paso anterior con las correspondencias de referencia para cada escena.

## 4. Resultados

El último paso permite obtener una medida cuantitativa del error de un conjunto de correspondencias, calificando indirectamente la eficacia de los métodos de detección de características. Además se obtiene una clasificación de cada correspondencia evaluando si su error es mayor que una determinada cota ( 1 píxel ), obteniendo así correspondencias pertenecientes a la matriz fundamental calculada o lo contrario, correspondencias atípicas utilizando terminología estadística.

Para las pruebas se utilizan los métodos con parámetros estándar en los que no se es muy estricto pero sin permitir que los métodos generen una gran cantidad de características. Este equilibrio es un problema que siempre va a surgir al tener que establecer la sensibilidad de los detectores para que se acomode a los deseos en cuanto a tiempo de procesamiento, flexibilidad, robustez, etc.

También hay que considerar que al aumentar el número de correspondencias que se utilizan para calcular la matriz fundamental, la proporción del número de correspondencias correctas disminuye, y a medida que se incluyan más hipótesis de correspondencias el número de incorrectas que se incorporan crece de forma exponencial.

### 4.1.1. Definición error de reproyección

En el error de reproyección de una matriz fundamental se pueden elegir diversas opciones ( mirar [12] ), se ha elegido la definición que se basa en la distancia que hay entre un punto y su línea epipolar para que tenga un significado geométrico que minimiza así las alteraciones que se realicen a la imagen al ser esta matriz un argumento en el proceso de rectificación posterior. Una línea  $l$  pasa por un punto  $m = (u, v)^T$  es descrita por la ecuación  $au + bv + c = 0$ , entonces  $l = [a, b, c]^T$ , que es un vector en 3D homogéneo por lo que  $l^T m = 0 = m^T l$ , ecuación que se puede multiplicar por cualquier escalar distinto de cero sin ser afectada. La distancia desde  $m$  a  $l$  será igual a :  $d(m, l) = \frac{au+bv+c}{\sqrt{a^2+b^2}}$  ; el error de reproyección, cantidad a minimizar, con ambas imágenes jugando el mismo papel respecto de la matriz fundamental sera :

$$\sum_i (d^2(m'_i F m_i) + d^2(m'_i F^T m_i))$$

Se compara el error de reproyección de la matriz de referencia sobre sus correspondencias con el error de reproyección de las matrices calculadas automáticamente pudiendo valorar la importancia relativa de los errores. En la cuadro 4.1 Avg significa la media del error sobre las correspondencias de referencia,



#### 4. Resultados

Escena \ Método	Harris		Harris-CNME		SURF		Referencia	
	Avg.	Dev.	Avg.	Dev.	Avg.	Dev.	Avg.	Dev.
Arch	96.0495	186.778	628.84	1344.31	138.226	314.22	0.419885	0.782094
Drive	76385.3	50770.2	3733.58	6859.55	8.42039	14.3209	2.55354	7.88287
Lab	10.8922	21.575	92.2438	205.26	4.66301	9.2451	0.563457	0.518607
Roof	258.983	534.574	25.8199	61.3615	3.66421	15.5262	2.67439	10.7416
Slate	5.24629	6.91642	1.45082	2.93642	0.66641	1.15358	0.0840964	0.139533
Yard	253.493	402.984	145.559	349.785	48.558	82.3075	0.796452	1.13143

Cuadro 4.1.: Error de reproyección de la matriz fundamental sobre las equivalencias de referencia para cada escena. Harris-CNME se refiere a Correlación Normalizada Multi Escala.

y Dev la desviación típica, en este cuadro se muestran los errores para los tres métodos de detección de características estudiados y las medidas de referencia.

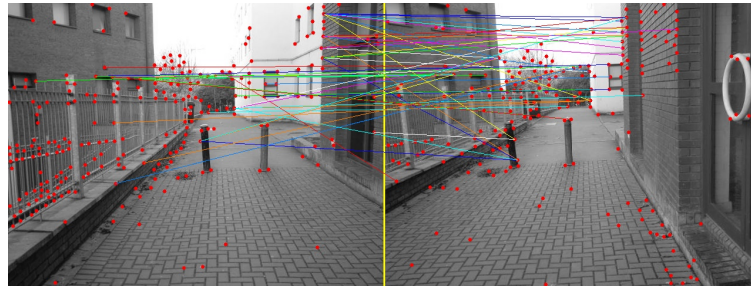
##### 4.1.2. Comentarios

En el apéndice A.6 están imágenes con los resultados de correspondencias para los tres métodos probados.

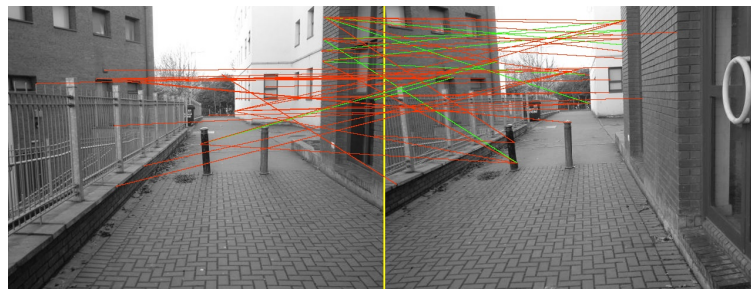
Con los datos se puede apreciar que el establecimiento de correspondencias basados en el método de detección por Harris y comparación por correlación normalizada no funciona, solo en la escena de Slate y Lab da resultados razonables debido en gran medida a la claridad de las escena y de sus esquinas, corroborando lo que se ve claramente en las imágenes de la figura 4.1.1 y 4.1.2.

La correlación normalizada solo funciona en los casos en los que la transformación entre las imágenes es muy cercana a una isotropía y cuando las estructuras de la imagen están bien definidas y poco repetidas, Ni siquiera la inclusión de diferentes escalas mejora los resultados. En Arch, Drive y Lab de hecho los empeora, son dos escenas que tienen varias estructuras repetidas, como ventanas y ramas de árbol, que alcanzan mayor correlación. Además al incluir las escalas se aumenta la probabilidad de que correspondencias basadas en parches similares pero no correspondientes alcancen el umbral de selección, y también disminuye la probabilidad de crear un correcto subconjunto mínimo de correspondencias (en RANSAC) al calcular la matriz fundamental. El buen resultado que presenta en la escena Slate es debido a que es una imagen en la que no hay estructuras repetidas, y las características tienen una clara localización y orientación. En la figura 4.1.1 se puede observar que prácticamente todas las correspondencias seleccionadas son erróneas y que la matriz fundamental también se calcula a partir de equivalencias equivocadas de la realidad.

#### 4. Resultados



(a) Selección de correspondencias de características Harris+CN.

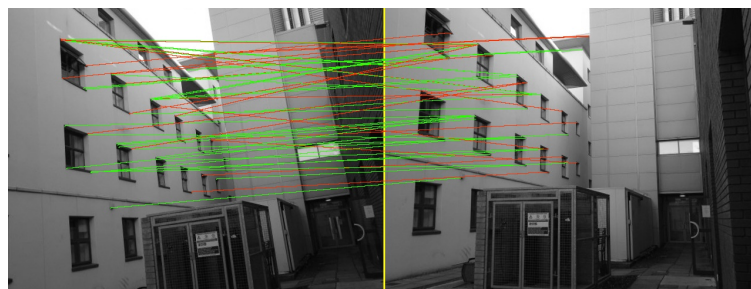


(b) Evaluación de las correspondencias.

Figura 4.1.1.: Evaluación en imágenes de las correspondencias obtenidas con Harris+CN de la escena Drive.



(a) Selección de correspondencias.



(b) Evaluación de las correspondencias.

Figura 4.1.2.: Evaluación en imágenes de las correspondencias obtenidas con Harris+CNME de la escena Arch.

## 4. Resultados

La correlación normalizada, aunque en principio se pensara que si es un buen método de correlación, lo cierto es que los resultados no es así, el ruido o pequeño desfase en el alineamiento de los parches afectan considerablemente a esta relación. Además la normalización de los valores hace que estructuras que espacialmente sean similares, con ligeras variaciones, pero que en el dominio de los valores si sean muy similares puedan alcanzar mayores correlaciones. Esto se puede apreciar en que hay ciertas estructuras que pueden alcanzar grandes cotas de correlación con una gran cantidad de parches, la multi escala solo aumenta el numero de correspondencias altas que tiene estas características. En la figura 4.1.2 fijarse en la esquina de la ventana más superior-izquierda de la imagen izquierda tiene correspondencias con varias esquinas de ventanas de la imagen derecha.

Sin embargo, SURF se acerca bastante a los datos de referencia fallando en la escena de Arch, así como teniendo un gran error en la escena de Yard. Este método destaca como una solución flexible, robusta, y capaz de discriminar en imágenes con alta repetición de características como la escena de Drive o la de Arch, o con poca presencia de características como Roof. Un caso curioso es la escena de Arch, ver figura 4.1.3, en la que a pesar de que visualmente aprecian numerosas correspondencias correctamente, alcanza un alto error respecto de nuestro conjunto de correspondencias de referencia. Esto es debido a que la distribución de las correspondencias sobre las imágenes influye en la matriz fundamental, en esta escena se establecen correctamente pero casi todas están sobre las ventanas de la fachada sobre la izquierda de la imagen. Esto hace que las correspondencias de referencia sobre la parte de la derecha contribuyan en gran medida al error de la matriz fundamental calculada.

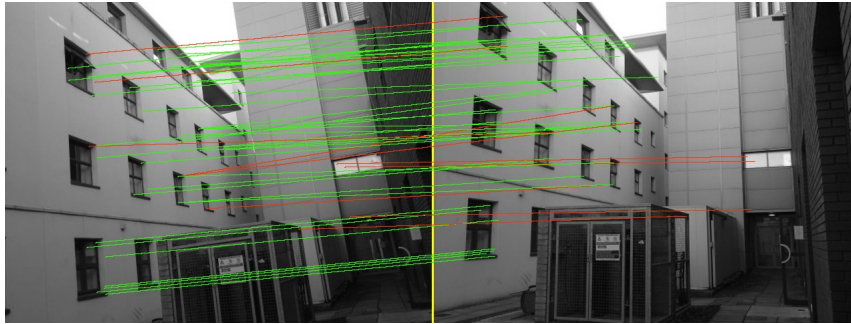
Obteniendo estos resultados se interpreta que los descriptores que capturan la tendencia general de los datos, como hace SURF con wavelets, obtendrán mejores resultados en la mayoría de los casos ya que reducen los efectos del ruido de las imágenes.. Por otra parte la detección precisa que puede hacer a través de las escalas aumenta la precisión en la localización de los puntos interesantes.

### 4.2. Medición de las distorsiones provocadas por la rectificación

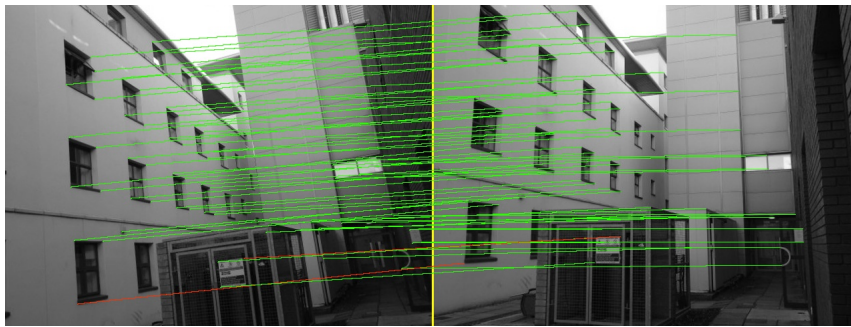
En el apéndiceA.7 están las imágenes producidas por los diferentes métodos de rectificación para poder compararlos visualmente..

Las homografías de rectificación se evalúan independientemente de la eficacia de los detectores de características utilizando las matrices fundamentales calculadas con las correspondencias establecidas

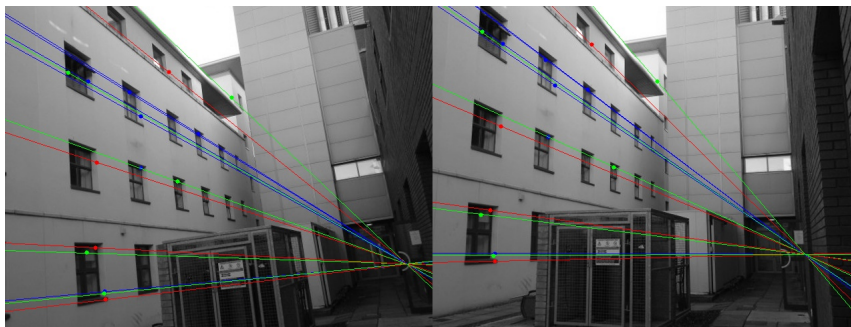
#### 4. Resultados



(a) Correspondencias con características SURF.



(b) Correspondencias establecidas manualmente como referencia.



(c) líneas epipolares de la matriz fundamental obtenida con correspondencias de SURF.

Figura 4.1.3.: Evaluación de las correspondencias en la escena Arch con SURF, en verde aceptadas, en rojo rechazadas.

#### 4. Resultados

manualmente, de esta forma esta garantizado que los datos con de los que se parte son correctos.

Para evaluar una rectificación se consideran varias cosas. Por un lado la precisión de la rectificación sobre una equivalencia se define como la distancia en su coordenada Y (ya que la rectificación “garantiza” el alineamiento de las filas de la imagen.), entonces la precisión general de una rectificación es calculada sobre el conjunto de equivalencias de referencia y esta definida como:

$$E_r = \|(Hm)_2 - (H'm')_2\|$$

Donde  $E_r$  es el error de rectificación,  $H$  y  $H'$  son las matrices de rectificación y  $m$  y  $m'$  son los puntos en cada imagen. El subíndice 2 de los paréntesis expresa que solo se opera con la componente Y.

Por otro lado se consideran las distorsiones introducidas en relación con los conceptos de perpendicularidad y de relación de aspecto en las imágenes modificadas, una rectificación será mejor en tanto que más se aproxime a los valores deseables de estas mediciones.

La perpendicularidad, su valor ideal es  $90^\circ$  grados, definido con los vectores que dividen a la imagen por la mitad en cada eje, ( ver 3.3.1) tal que:

$$E_p = \arccos\left(\frac{X \cdot Y}{|X||Y|}\right)$$

Relación de aspecto, cuyo valor ideal es 1, es una medición sobre la relación entre el tamaño del eje X y del eje Y, por tanto con las cuatro esquinas del cuadrado de la imagen ( rectificada ),  $p_1 = (w/2, 0, 1)^T$ ,  $p_2 = (w, h/2, 1)^T$ ,  $p_3 = (w/2, h, 1)^T$ ,  $p_4 = (0, h/2, 1)^T$ , ( como antes  $w$  es el ancho y  $h$  es la altura ) se establecen los vectores que dividan la imagen en 2 tal que:

$$X = (x_u, y_v)^T = \hat{p}_2 - \hat{p}_4$$

$$Y = (x_u, y_v)^T = \hat{p}_1 - \hat{p}_3$$

La relación de aspecto será:

$$E_a = \left(\frac{X^T X}{Y^T Y}\right)^{\frac{1}{2}}$$

#### 4. Resultados

(a) Medición de la rectificación para el método de Hartley.

Escena \ Método	Hartley					
	$E_r$		$E_p$		$E_a$	
	Avg	Dev	Izda	Dcha	Izda	Dcha
Arch	0.346274	0.299964	93.8882	90	1.06304	1.00032
Drive	0.602625	1.01271	109.234	90	1.40439	0.998931
Lab	0.454675	0.282278	82.1333	90	0.861112	0.987612
Roof	0.608841	0.907541	127.737	90	2.04785	1.00185
Slate	0.148417	0.13944	91.9485	90	1.03531	1.00251
Yard	0.478703	0.410754	105.824	90	1.30993	1.00001

(b) Medición de la rectificación con la modificación sobre perpendicularidad y relación de aspecto.

Escena \ Método	Hartley+Torsión					
	$E_r$		$E_p$		$E_a$	
	Avg	Dev	Izda	Dcha	Izda	Dcha
Arch	0.346274	0.299964	92.8475	92.154	1.04543	1.03762
Drive	0.602625	1.01271	97.8386	97.8742	1.17195	1.14516
Lab	0.454675	0.282278	97.0601	98.4011	1.10428	1.1402
Roof	0.608841	0.907541	90.3398	93.2002	1.01243	1.05834
Slate	0.148417	0.13944	88.6981	88.5094	0.980671	0.978468
Yard	0.478703	0.410754	89.7785	90.5015	0.998606	1.00848

(c) Medición de la rectificación del método de Whelan.

Escena \ Método	Whelan					
	$E_r$		$E_p$		$E_a$	
	Avg	Dev	Izda	Dcha	Izda	Dcha
Arch	0.327335	0.283266	91.2215	90.2156	1.01606	1.00357
Drive	0.493783	0.830005	87.868	89.7786	0.988321	0.996432
Lab	0.394685	0.244869	87.1544	87.2495	0.946659	0.955818
Roof	0.554754	0.828066	89.9559	90.7279	1.0045	1.012
Slate	0.160596	0.149994	88.5889	88.3163	0.97157	0.976037
Yard	0.472881	0.405743	89.7561	90.0424	0.998224	1.00071

Cuadro 4.2.: Medición de las rectificaciones.

## 4. Resultados

### 4.2.1. Comentarios

En la cuadro 4.2 están los errores para cada uno de los métodos analizados, en la figura 4.2.1 se muestran estos métodos aplicados a la escena de Roof, a la izquierda la imagen, a la derecha una malla de líneas para apreciar mejor la distorsión provocada. Respecto del error de rectificación, se observa como el método Hartley y la modificación con la matriz de torsión (que mantiene la perpendicularidad y la relación de aspecto) no modifica la relación de las líneas epipolares al no modificar la componente Y de los puntos. Además se ve como en la escena Roof estos métodos son los que minimizan más este error, sin embargo, en el caso de Hartley introduce gran distorsión en la imagen, mientras la modificación permite mantener este error mejorando el aspecto visual de la imagen. En las escenas de Drive, Roof y Yard son las que visualmente salen más torcidas con el método de Hartley y en las que se puede apreciar una mejora mayor en este aspecto. La transformación casi “rígida” que hace el método de Hartley de una de las imágenes se ve reflejado en los datos de las imágenes Dcha. donde su perpendicularidad y su relación de aspecto son los valores ideales  $90^\circ$  grados y 1 respectivamente.

La ventaja del método de Whelan es que se basa únicamente en la matriz fundamental por lo que, si esta es correcta, consigue una rectificación que mantiene sus propiedades afines de forma local manteniendo así la apariencia visual de la imagen. Como Whelan explica, su método es superior al considerar el criterio de Zhang sobre toda la imagen. Tiene la desventaja en que no es aplicable en el caso de que alguno de los epipolos esté cerca de la imagen, en este caso las distorsiones introducidas harán a la imagen irreconocible, pero este no suele ser el caso de sistemas estéreo.

Unas pocas correspondencias pueden afectar de manera importante a la matriz fundamental, esto se puede comprobar en el cuadro 4.2.2, en la columna de la izquierda, las correspondencias marcadas en rojo hacen que las líneas epipolares varíen significativamente, por tanto las imágenes rectificadas se verán afectadas en la medida en la que lo sean las líneas epipolares.

Un aspecto que no medido aquí es la disparidad que tienen los puntos de las correspondencias, es decir, la diferencia en la coordenada X, este aspecto tiene sentido tratarlo con un sistema estéreo calibrado o bajo la premisa de matrices de parámetros intrínsecos de ambas cámaras iguales. En el

#### 4. Resultados

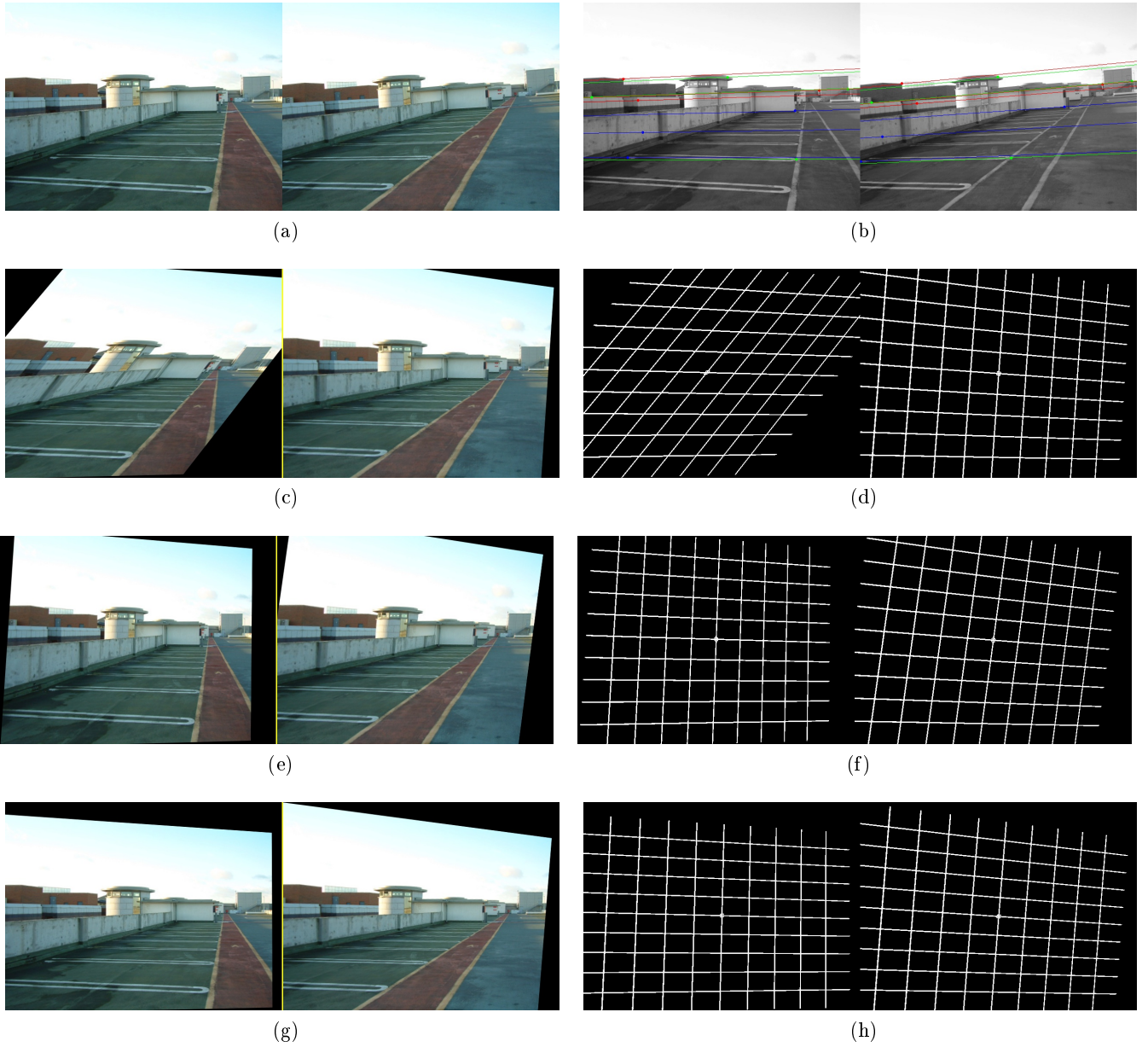


Figura 4.2.1.: Resultado de aplicar diferentes métodos de rectificación a la pareja de imágenes original y con las líneas epipolares (a) y (b), de la escena Roof. (c) y (d) rectificación de Hartley, (e) y (f) Hartley con corrección de torsión, (g) y (h) rectificación de Whelan.



#### 4. Resultados

desarrollo de varias soluciones se comprobó que hay que alcanzar una solución de compromiso entre minimizar esta diferencia y mantener la apariencia visual de la imagen.

Estas imágenes rectificadas sirven de entrada para el cálculo de las imágenes de disparidad y, conocida la distancia focal, de la profundidad, en estos algoritmos se tendrá que lidiar entre una mayor diferencia entre escalas pero con menor variación en la ortogonalidad de las direcciones, o con menor diferencia entre la posición de los puntos y mayor variación en las direcciones. Recordar que en el apéndice A.7 se pueden comparar visualmente las diferentes alternativas.

#### 4. Resultados

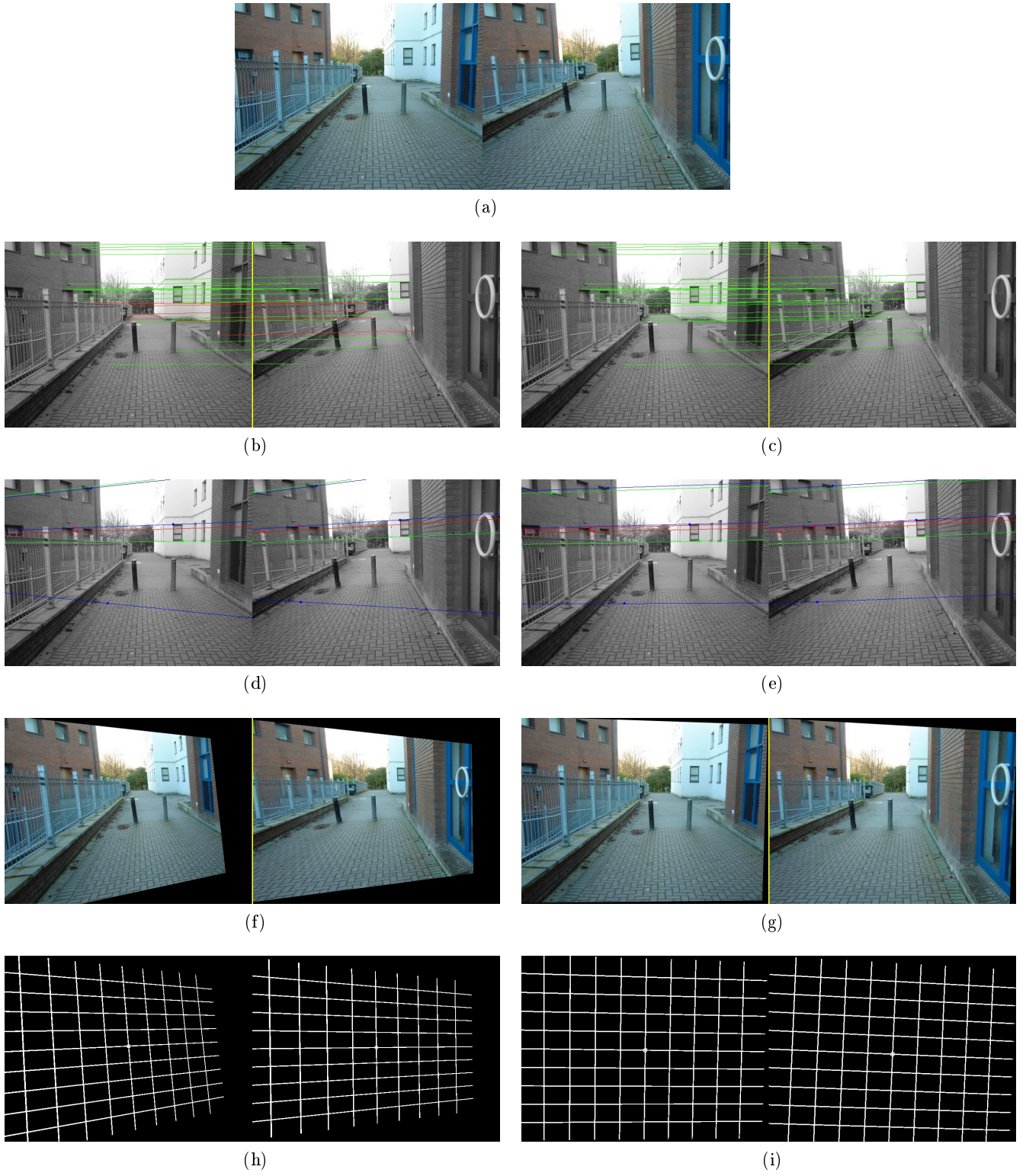


Figura 4.2.2.: Fragilidad de la matriz fundamental provocando diferentes imágenes rectificadas con el método de Whelan.

## 4. Resultados

### 4.3. Proceso Completo

En esta sección vamos a ver el proceso completo aplicado a la escena Drive ilustrado en imágenes.



Figura 4.3.1.: Pareja de imágenes originales.

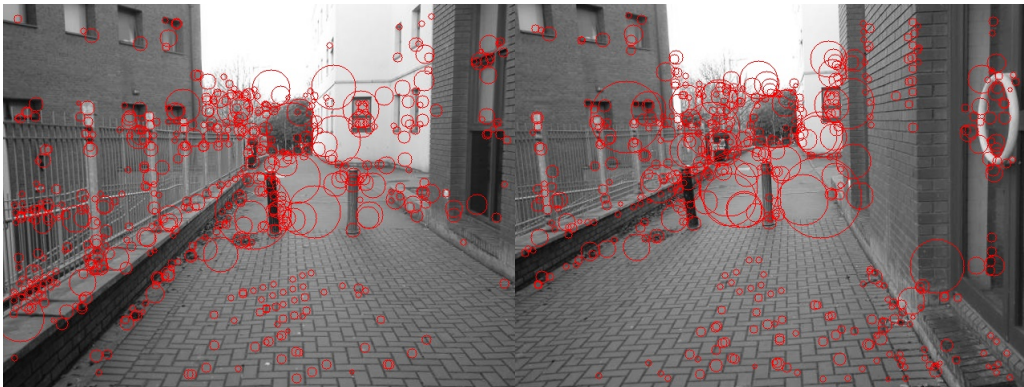


Figura 4.3.2.: Detección de características SURF.



#### 4. Resultados

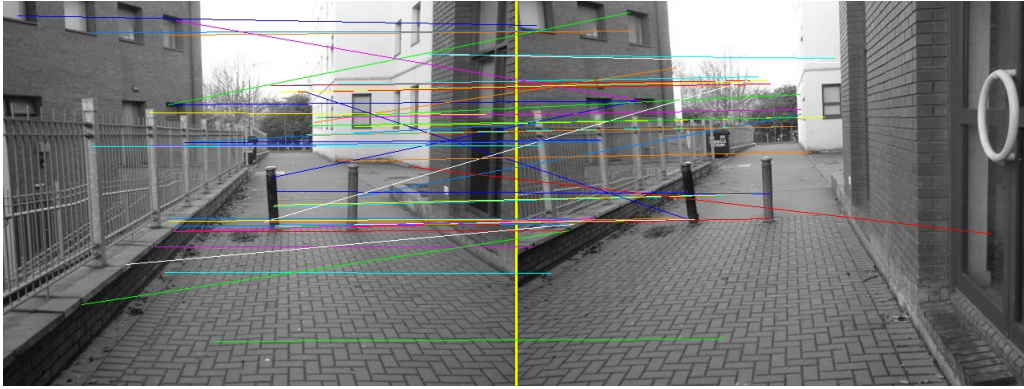
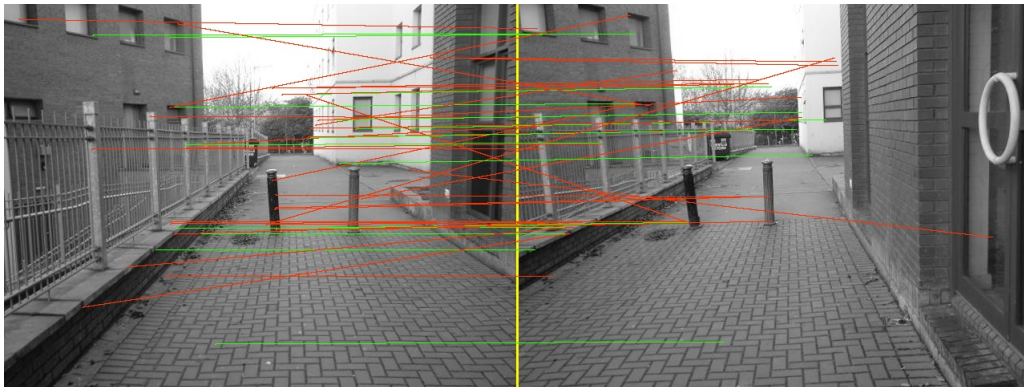
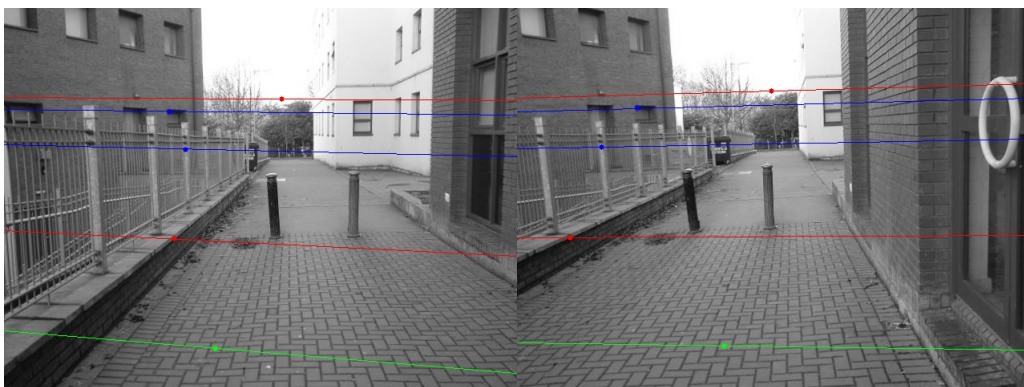


Figura 4.3.3.: Selección de equivalencias entre imágenes.



(a) Evaluación de las equivalencias según la matriz fundamental, correctas en verde, incorrectas en rojo..



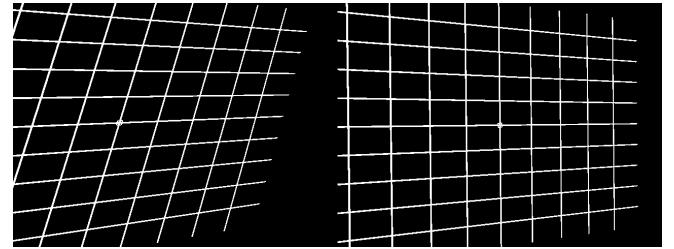
(b) líneas Epipolares.

Figura 4.3.4.: Cálculo de la matriz fundamental

#### 4. Resultados



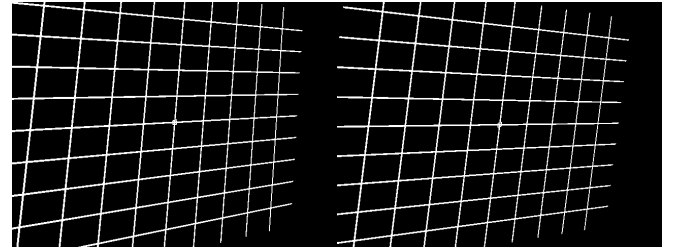
(a)



(b)



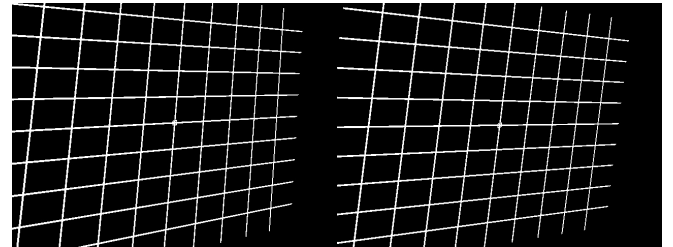
(c)



(d)



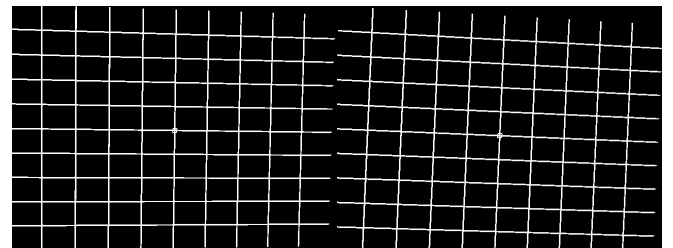
(e)



(f)



(g)



(h)

Figura 4.3.5.: Rectificación de imágenes.

## 5. Conclusiones

En este trabajo de investigación se han estudiado los pasos necesarios para mejorar la coherencia visual entre 1 par de imágenes en un sistema estéreo de cámaras, en resumen:

- Se han visto los enfoques más exitosos para conseguir encontrar correspondencias entre imágenes, viendo cuales son los puntos fuertes de SURF y como la importancia del descriptor que margine las características es esencial.
- Se han analizado diversos métodos de rectificación y como afecta cada paso al resultado final, llegando a una solución en la que se consiguen rectificar las imágenes sin apenas modificar las propiedades afines de las imágenes originales si se parte de una matriz fundamental correcta..

Para este estudio se desarrollo el software necesario para las pruebas de las diferentes soluciones, pudiendo ser integrado en algún sistema como por ejemplo el Ivvi desarrollado en el departamento de Automática.

Este estudio es solo un primer paso para mejorar los sistemas estéreo basados únicamente en información visual pudiendo ser objeto de ampliación y mejora.

### Futura Investigación

Hay varias líneas de investigación que seguro mejoraran los resultados entre las se proponen:

Sobre la detección de características, es probable que un sistema combinado en el que se utilice tanto SURF como Harris de mejores resultados ya que son métodos complementarios.

Se puede investigar sobre el diseño de un descriptor más avanzado para las esquinas de Harris, con selección del tamaño del parche, quizás aplicar técnicas de estadística robusta sobre los parches o algún sistema como PHOG (Pyramid Histogram of Oriented Gradients, ver [20]) utilizado en el área de reconocimiento de objetos.

## 5. Conclusiones

Además restringir la acumulación de características en determinadas zonas de la imagen mejorará los resultados, en la detección sobre varias escalas evitaría la duplicación de puntos que se utilizan para buscar correspondencias.

Sobre el cálculo de la matriz fundamental, es posible que se pueda mejorar el esquema RANSAC sobre el conjunto de hipótesis de correspondencias, en [21] se utiliza un sistema combinado con la transformada de Hough. Este esquema probablemente sea capaz de encontrar la matriz fundamental en un conjunto de correspondencias más contaminado y por tanto se podrá aumentar el tamaño del conjunto utilizado aumentando también la precisión de la matriz.

Sobre la rectificación de imágenes, otras soluciones pueden ser aplicadas, pero básicamente son soluciones que aplican restricciones que puedan ajustarse mejor a las características de un sistema concreto. El algoritmo que se presenta, publicado por Whelan, es un método preciso y flexible, teniendo la precaución de alejar los epipolos de las imágenes.

# A. Apéndices

## A.1. Presupuesto

En este estudio únicamente se ha utilizado equipamiento básico de oficina, un equipo informático con conexión a Internet. Además el uso de las librerías OpenCv, de código abierto y gratuitas, ha permitido aligerar la carga de desarrollo necesaria reduciendo así el tiempo de ejecución y por tanto, su coste.

El tiempo empleado lo seccionamos de acuerdo con las partes de la memoria ∴

- Fase 1:
  - Estudio de características, 25 horas
  - Implementación, depuración, 30 horas
- Fase 2
  - Estudio de geometría, 30 horas
  - Implementación, 50 horas
- Fase 3
  - Realización memoria, 20 horas
- Total horas = *155 horas*



## A. Apéndices

Suponiendo unos costes de :

- Coste neto asalariado por hora = 25 €
- IRPF = 20 %
- Seguridad Social empresa = 30 %
- Seguridad Social asalariado = 7 %

El coste total del estudio seria de :

$$C_A = 155 \cdot 25 = 3875 \text{ €}$$

$$C_{IRPF} = C_A \cdot 0,2 = 775 \text{ €}$$

$$C_{SSE} = C_A \cdot 0,3 = 1162,5 \text{ €}$$

$$C_{SSA} = C_A \cdot 0,07 = 271,25 \text{ €}$$

$$C_{bruto\ total} = C_A + C_{IRPF} + C_{SSE} + C_{SSA} = \underline{6083,75 \text{ €}}$$

## A.2. Imagen integral

La idea de la imagen integral consiste en calcular la suma acumulada de los valores de la imagen a lo largo de los ejes de coordenadas, de esta forma se puede calcular la suma entre cualquier intervalo de una línea con 2 accesos a memoria, y la de cualquier cuadrado, alineado sobre los ejes, con 4 accesos. Esta imagen acelera el cálculo de wavelets de Haar, por lo que es utilizado en la extracción de las características de SURF.

La imagen integral se define en un punto de la imagen como:

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y')$$

Donde  $ii(x, y)$  es el valor de la imagen integral,  $i(x', y')$  es el valor de la imagen original. Usando la imagen integral se puede calcular la suma de los píxeles del cuadrado D de la figura A.2.1 con 4 accesos, a los cuatro puntos señalados con 1,2,3,4 con la ecuación  $D = 4 - 3 - 2 + 1$ .

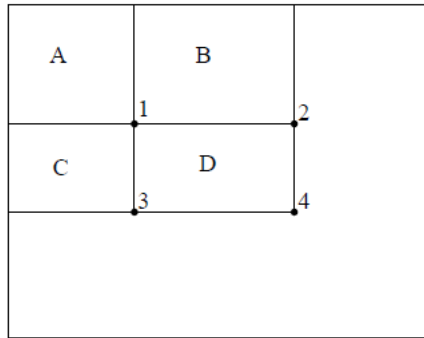


Figura A.2.1.: Imagen integral

### A.3. Homografía

En visión artificial una *homografía* es una relación proyectiva entre dos planos, por ejemplo la relación que existe entre una superficie plana del mundo real y su imagen capturada por una cámara ( con el modelo ideal, sin desplazamientos por imperfecciones ni distorsiones ) es descrita por una homografía. Esto puede ser demostrado, supongamos que la relación entre el punto  $Q = [X \ Y \ Z \ 1]^T$  en la escena y el punto  $q = [x \ y \ 1]^T$  en la imagen es una proyección y puede ser expresada como  $q = PQ$ . Podemos elegir el sistema de referencia de tal forma que el plano  $XY$  coincida con el plano de la escena  $\Pi$  en el que se encuentra  $Q$ , ver figura A.3.1, de esta forma conseguimos que este punto tenga la componente  $Z$  igual a cero, entonces si denotamos las columnas de  $P$  como  $p_n$  podemos expresar la proyección como :

$$q = PQ = [p_1 \ p_2 \ p_3 \ p_4] \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = [p_1 \ p_2 \ p_4] \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

Por lo que la relación entre los puntos  $Q_\Pi = [X \ Y \ 1]^T$  del plano  $\Pi$  y sus proyecciones en el plano imagen es una homografía que cumple  $q = HQ_\Pi$  siendo  $H$  una matriz 3x3 .

En el proceso de calibración la homografía es la herramienta básica que se utiliza para el cálculo de los parámetros intrínsecos que se están buscando. La homografía puede ser descompuesta así :

$$q = sM [r_1 \ r_2 \ t]$$

Donde  $s$  es un factor de escala,  $M$  es la matriz de los parámetros intrínsecos de la cámara, la matriz entre corchetes representarían columnas de una matriz donde  $r_1$  y  $r_2$  serían columnas de rotación y  $t$  columna de desplazamiento.

La homografía aparece también en la relación entre los puntos de dos imágenes no calibradas como matriz fundamental.

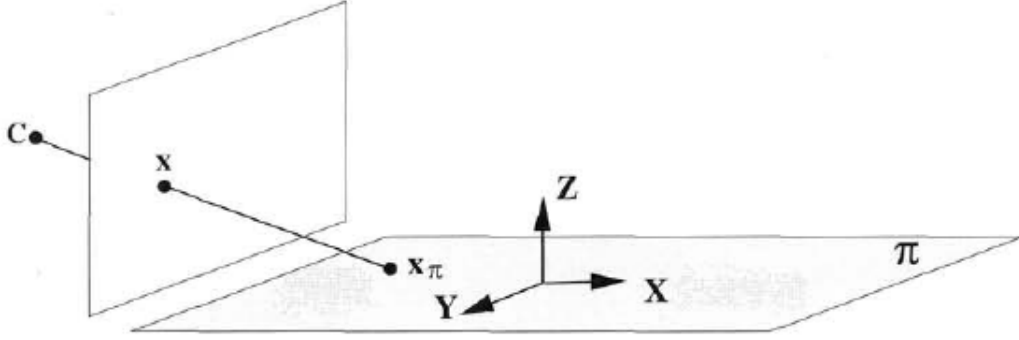


Figura A.3.1.: Elección sistema de referencia para el cálculo de la homografía que los relaciona.

#### A.4. Deducción de homografía pareja en Hartley

Esta es una deducción simplificada de la homografía pareja, para más detalles ver [17], el desarrollo requiere de alguna terminología previa.

Si  $A$  es una matriz cuadrada, la matriz de cofactores sera  $A^*$ , con las que se establecen las propiedades :  $A^*A = AA^* = \det(A)I$  , en concreto si  $A$  es invertible entonces  $A^* \approx (A^T)^{-1}$ .

Por otro lado, dado un vector  $t = (t_x, t_y, t_z)^T$  con el que podemos formar la matriz antisimétrica:

$$[t]_{\times} = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}$$

Para cualquier vector distinto de cero  $t$ , la matriz  $[t]_{\times}$  tiene rango 2 cumpliéndose además que  $t^T[t]_{\times} = [t]_{\times}t = 0$  y que cualquier otro vector anulado por  $[t]_{\times}$  es un múltiplo escalar de  $t$ .

Para cualquier matriz  $M$  de tamaño  $3 \times 3$  , y cualquier vector  $t$  se establece que:

$$M^*[t]_{\times} = [Mt]_{\times}M$$

Si  $M$  es una matriz 3x3 no singular que representa una transformación proyectiva en  $\mathbb{P}^2$ , entonces  $M^*$  es la correspondiente transformación para líneas, es decir, si  $u_1$  y  $u_2$  forman la línea  $l$  , entonces  $Mu_1$  y  $Mu_2$  será la línea  $M^*l$  , en ecuación :

$$(Mu_1) \times (Mu_2) = M^*(u_1 \times u_2)$$

A parte de las propiedades de la matriz fundamental, esta se puede descomponer tal que:

## A. Apéndices

$$F = [e']_{\times} M = M^* [e]_{\times}$$

Esto supone que si tenemos una línea  $l$  en la primera imagen que pasa por el epipolo  $e$ , es decir, es una línea epipolar, entonces  $M^*l$  será la línea epipolar correspondiente en la segunda imagen

La propiedad de la matriz fundamental de  $Fe = 0 = F^T e'$  supone que la descomposición de esta matriz no es única, si tenemos 2 maneras diferentes  $F = S_1 M_1 = S_2 M_2$  donde  $S_1$  y  $S_2$  son matrices antisimétricas y  $M_1$  y  $M_2$  son matrices no singulares, entonces  $S_1 = S_2$  restringiendo además que si  $S_1 = [e']_{\times}$  entonces  $M_2 = (I + e'a^T)M_1$  para algún vector  $a$ . Esta restricción permite obtener la expresión para la homografía pareja .

Si  $u$  es un punto en  $J$ , entonces  $e \times u$  es la línea epipolar en  $J$ , y  $Fu$  es la línea epipolar en la segunda imagen  $J'$ , las homografías de rectificación son compatibles solo si  $H^*(e \times u) = H'^*Fu$ , como esto se debe cumplir para cualquier  $u$ , podemos escribir  $H^*[e]_{\times} = H'^*F = H'^*[e']_{\times}M$ , o expresado sobre los puntos  $[He]_{\times}H = [H'e']_{\times}H'M$  de donde puede deducir finalmente que la homografía pareja a  $H'$  tiene la forma  $H = (I - H'e'a^T)H'M$  para algún vector  $a$ .

## A. Apéndices

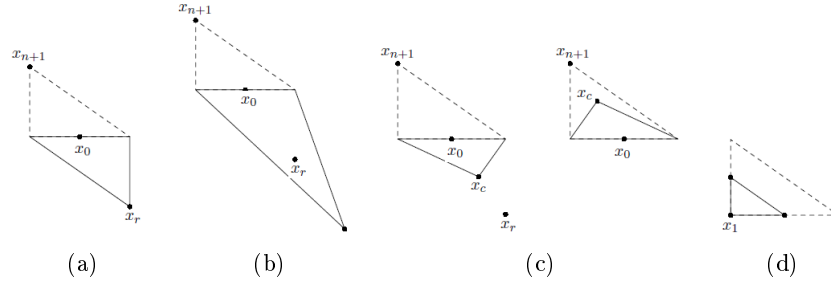


Figura A.5.1.: Fases del método simplex de optimización, a) reflexión, b) expansión, c) contracción, 2 casos posibles, d) reducción.

### A.5. Método Nelder-Mead simplex

Este método de optimización [15] no lineal se ha convertido en uno de los más utilizados desde su publicación en 1965, únicamente se basa en los valores de la función a optimizar, no utiliza derivadas, entrando en la categoría de método directo de búsqueda. En cada iteración se parte un simplex, una figura geométrica en  $n$  dimensiones que es la envoltura convexa de  $n+1$  vértices, en cada vértice se evalúa la función

Este método tiene 4 parámetros configurables: coeficientes de reflexión  $\alpha$ , de expansión  $\beta$ , de contracción  $\gamma$ , de reducción  $\delta$ .

- 1- Ordenar los vértices tal que  $f(x_1) \leq f(x_2) \leq \dots \leq f(x_{n+1})$
- 2- Calcular el centroide de los puntos menos el ultimo  $x_0 = \sum_{i=1}^n x_i / n$
- 3- Reflexión: calcular el punto de reflexión  $x_r = x_0 + \alpha(x_0 - x_{n+1})$ , evalúa  $f_r = f(x_r)$ , y si  $f_1 \leq f_r < f_n$ , reemplaza  $x_{n+1}$  por  $x_r$  y ves a la siguiente iteración, ( vas al paso 1 ).
- 4- Expansión: si  $f_r \leq f_1$  entonces calcula el punto de expansión  $x_e = x_0 + \beta(x_0 - x_{n+1})$ , evalúa  $f_e = f(x_e)$ , y si  $f_e \leq f_r$ , reemplaza  $x_{n+1}$  por  $x_e$  y ves a la siguiente iteración, si  $f_e \geq f_r$  reemplaza  $x_{n+1}$  por  $x_r$  y ves a la siguiente iteración, en otro caso continua
- 5- Contracción: debe cumplirse que  $f_n \leq f_r$ , calcula el punto de contracción  $x_c = x_{n+1} + \gamma(x_0 - x_{n+1})$ , evalúa  $f_c = f(x_c)$ , si  $f_c \leq f_{n+1}$ , reemplaza  $x_{n+1}$  por  $x_c$  y ves a la siguiente iteración, en otro caso continua.
- 6- Reducción: para todos los puntos menos el mejor,  $i = 2, \dots, n + 1$ , reemplaza los puntos tal que  $x_i = x_1 + \delta(x_i - x_1)$ .

## **A.6. Imágenes de Correspondencias**

## A. Apéndices

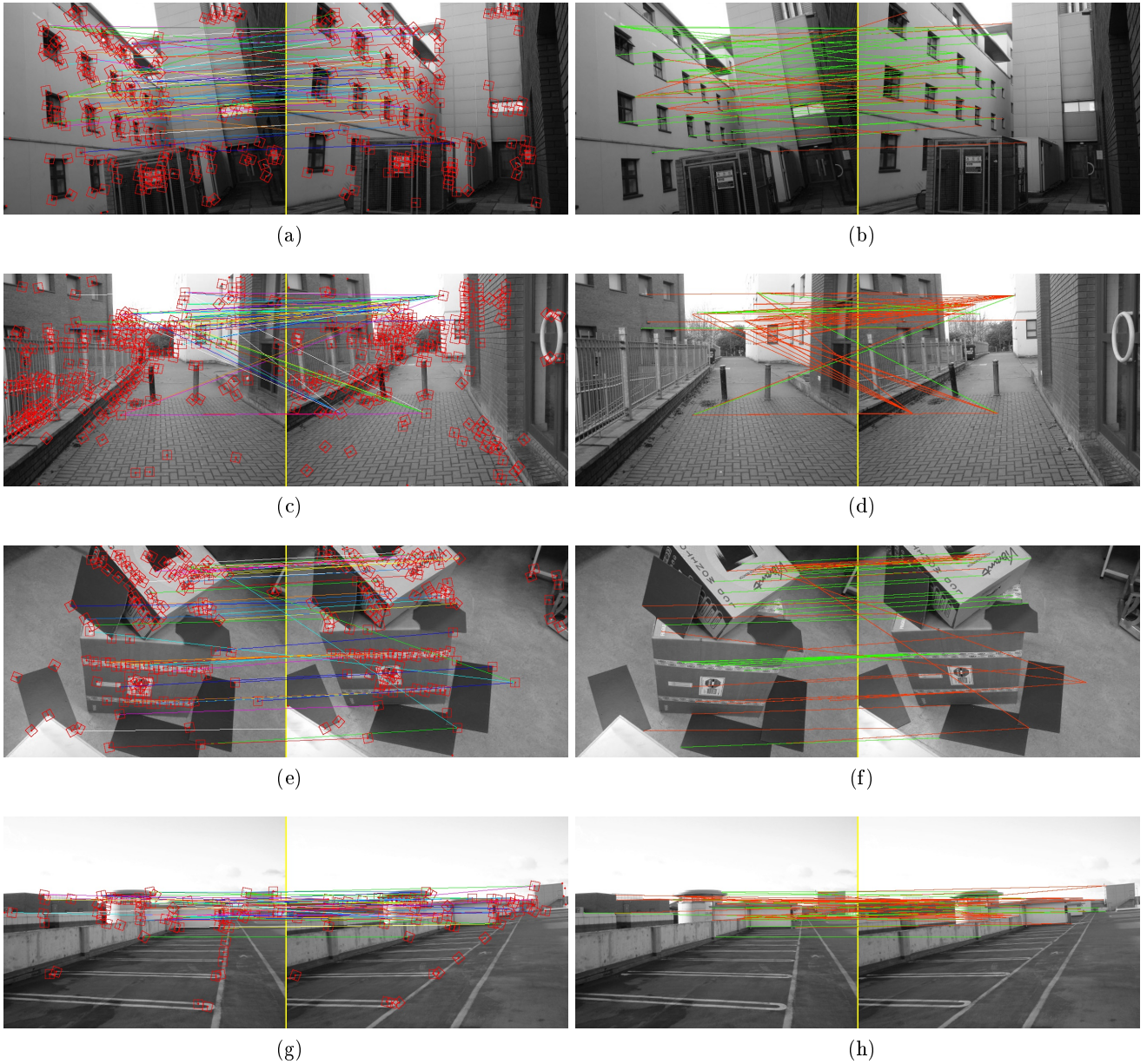


Figura A.6.1.: Imágenes de correspondencias de Harris+CN para (a y b) escena Arch, (c y d) escena Drive, (e y f) escena Lab, (g y h) escena Roof.



## A. Apéndices

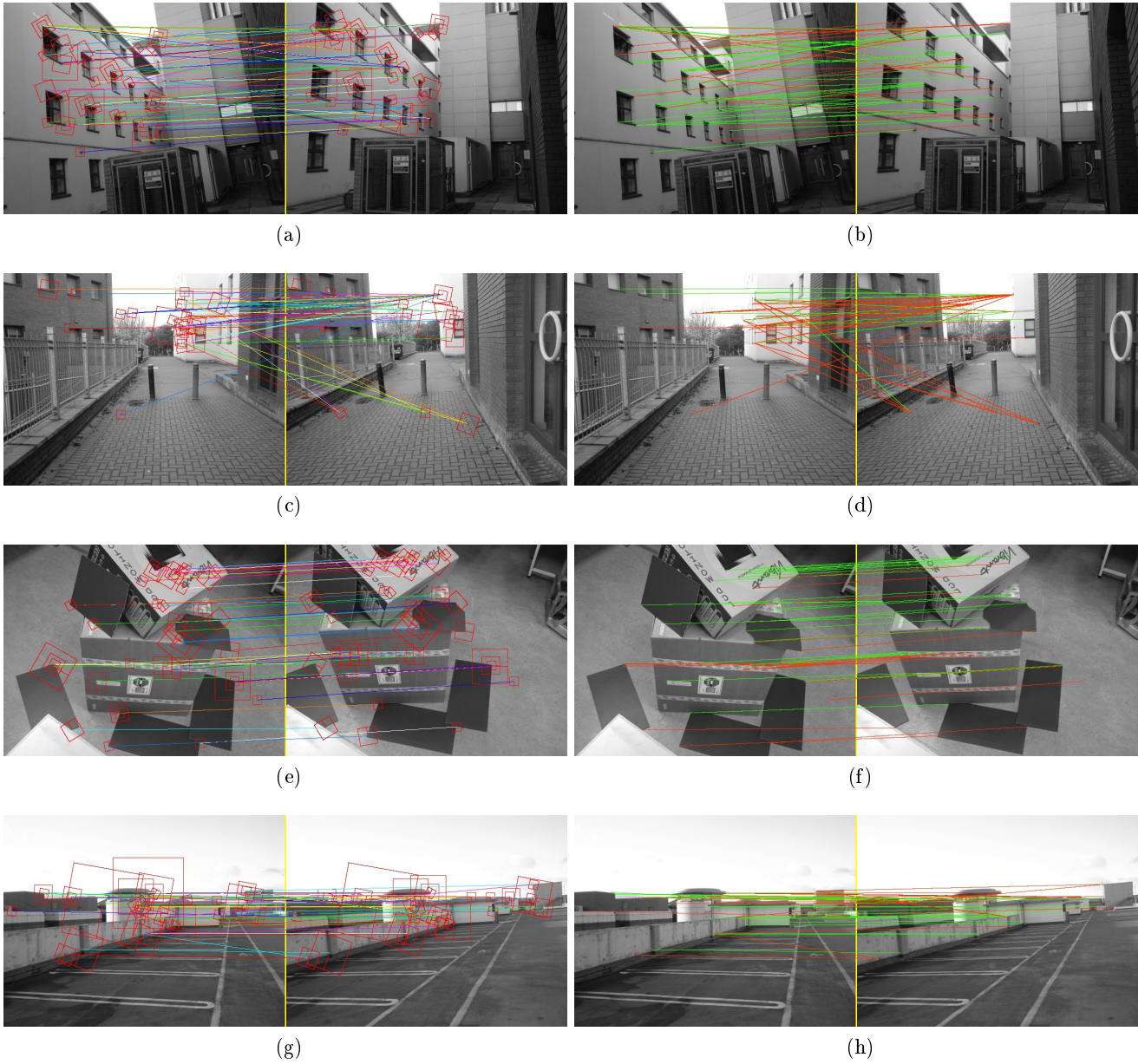


Figura A.6.2.: Imágenes de correspondencias de Harris+CNME para (a y b) escena Arch, (c y d) escena Drive, (e y f) escena Lab, (g y h) escena Roof.

## A. Apéndices

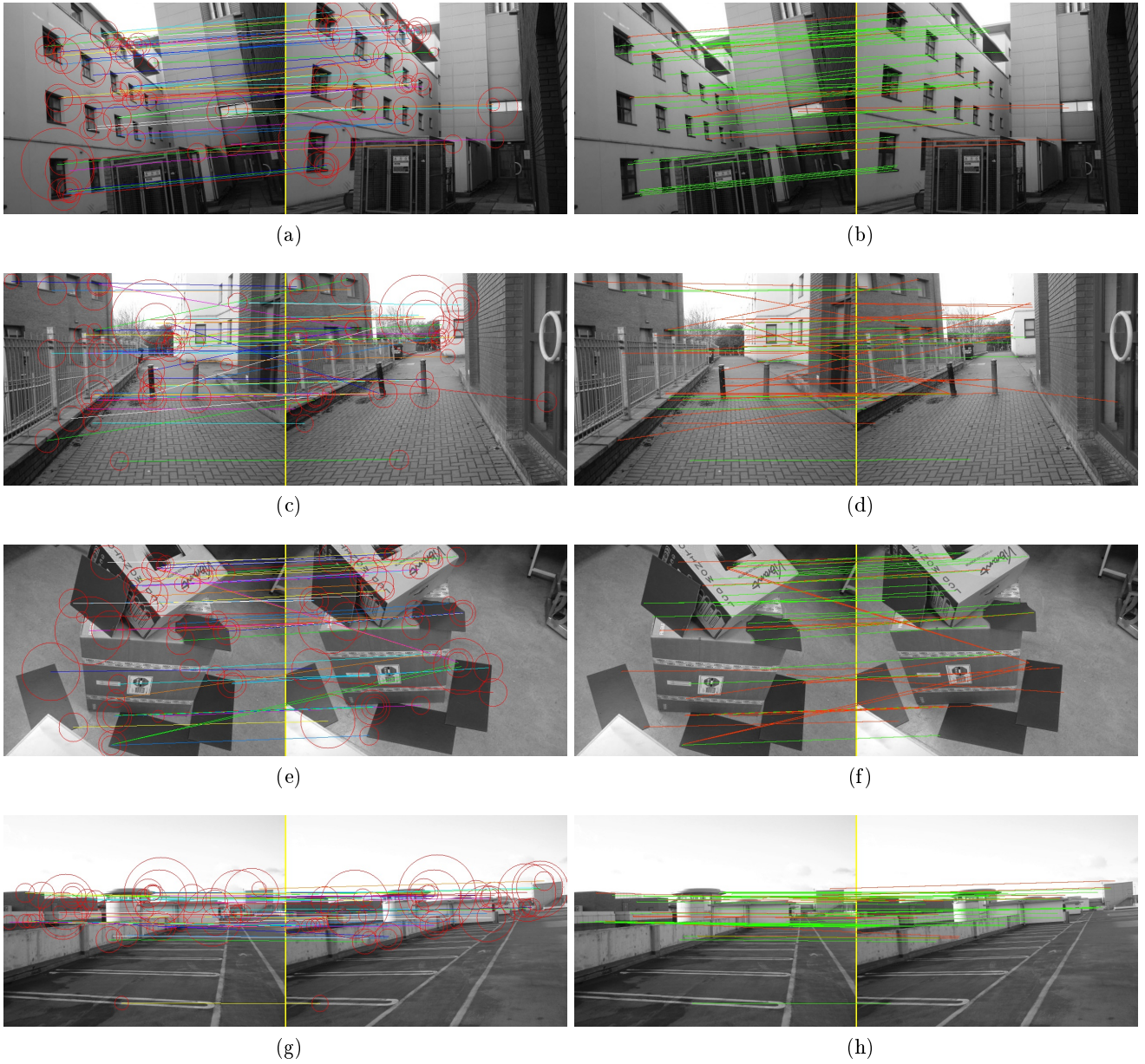


Figura A.6.3.: Imágenes de correspondencias de SURF para (a y b) escena Arch, (c y d) escena Drive, (e y f) escena Lab, (g y h) escena Roof.

## **A.7. Imágenes de Rectificaciones**

## A. Apéndices

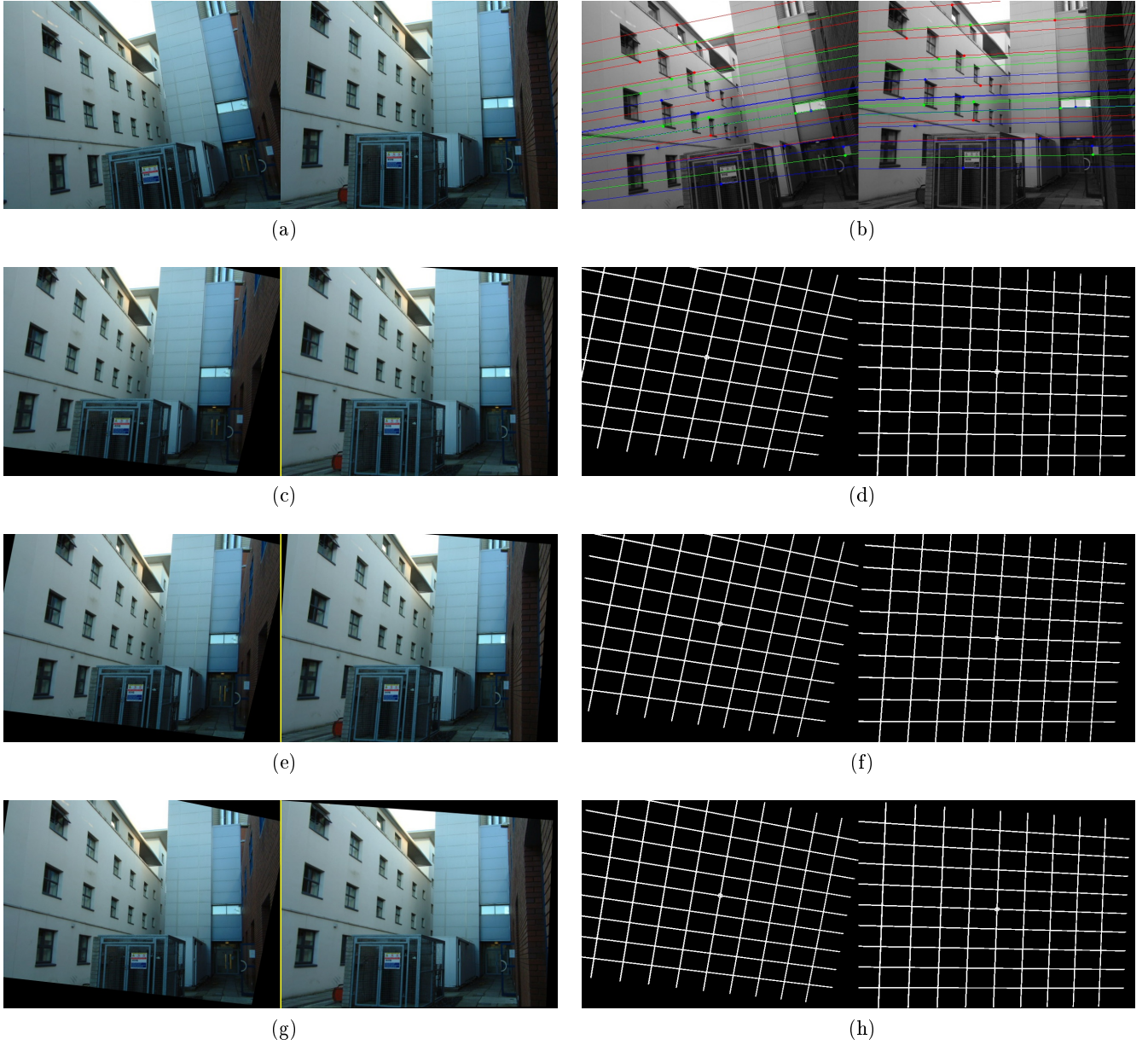


Figura A.7.1.: Rectificación de imágenes para la escena Arch, primera fila, original y original con líneas epipolares, segunda fila Hartley, tercera fila, Hartley + Zhang, cuarta fila Whelan .



## A. Apéndices

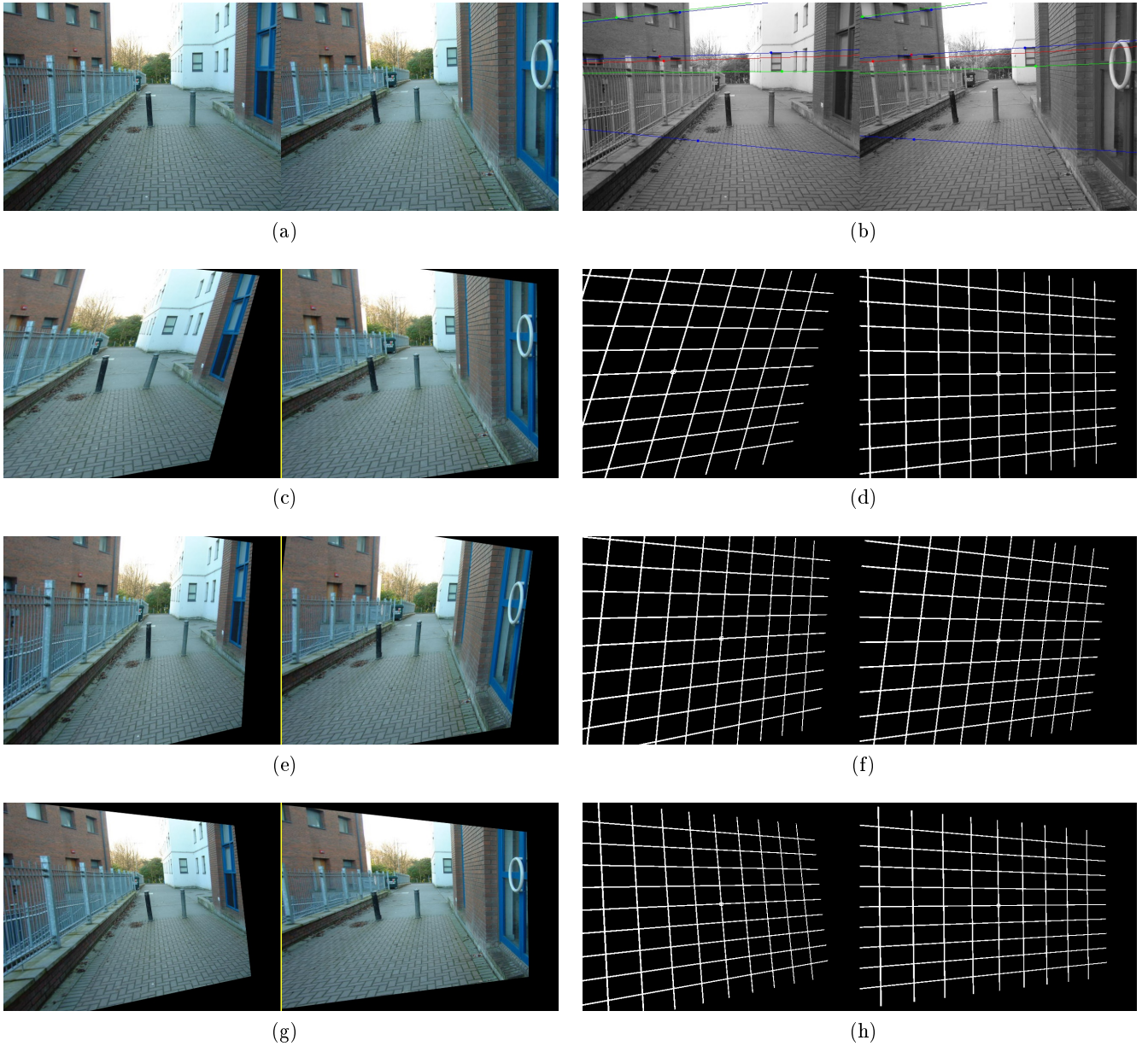
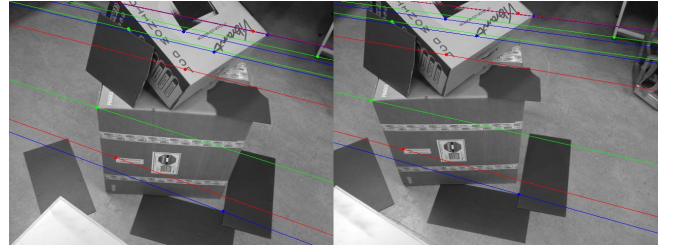


Figura A.7.2.: Rectificación de imágenes para la escena Drive, primera fila, original y original con líneas epipolares, segunda fila Hartley, tercera fila, Hartley + Zhang, cuarta fila Whelan.

## A. Apéndices



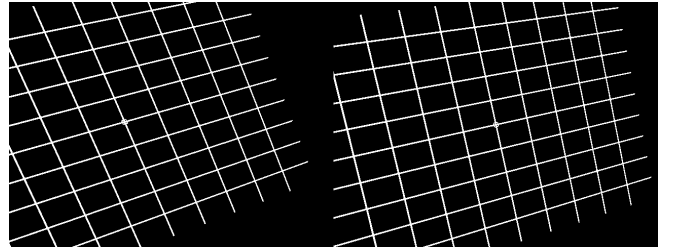
(a)



(b)



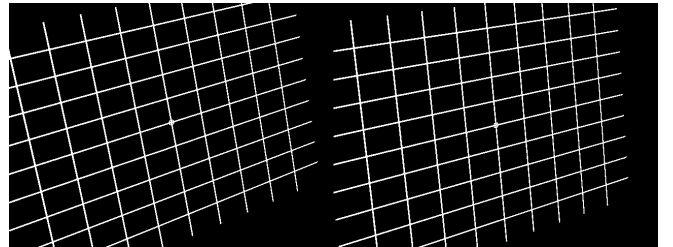
(c)



(d)



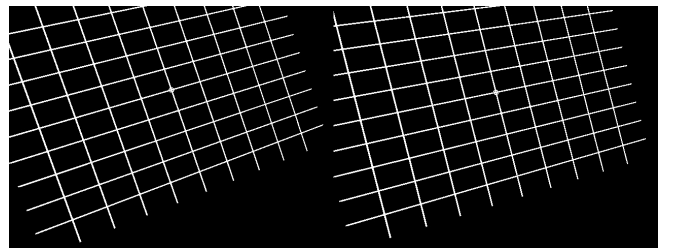
(e)



(f)



(g)



(h)

Figura A.7.3.: Rectificación de imágenes para la escena Lab, primera fila, original y original con líneas epipolares, segunda fila Hartley, tercera fila, Hartley + Zhang, cuarta fila Whelan .

## A. Apéndices

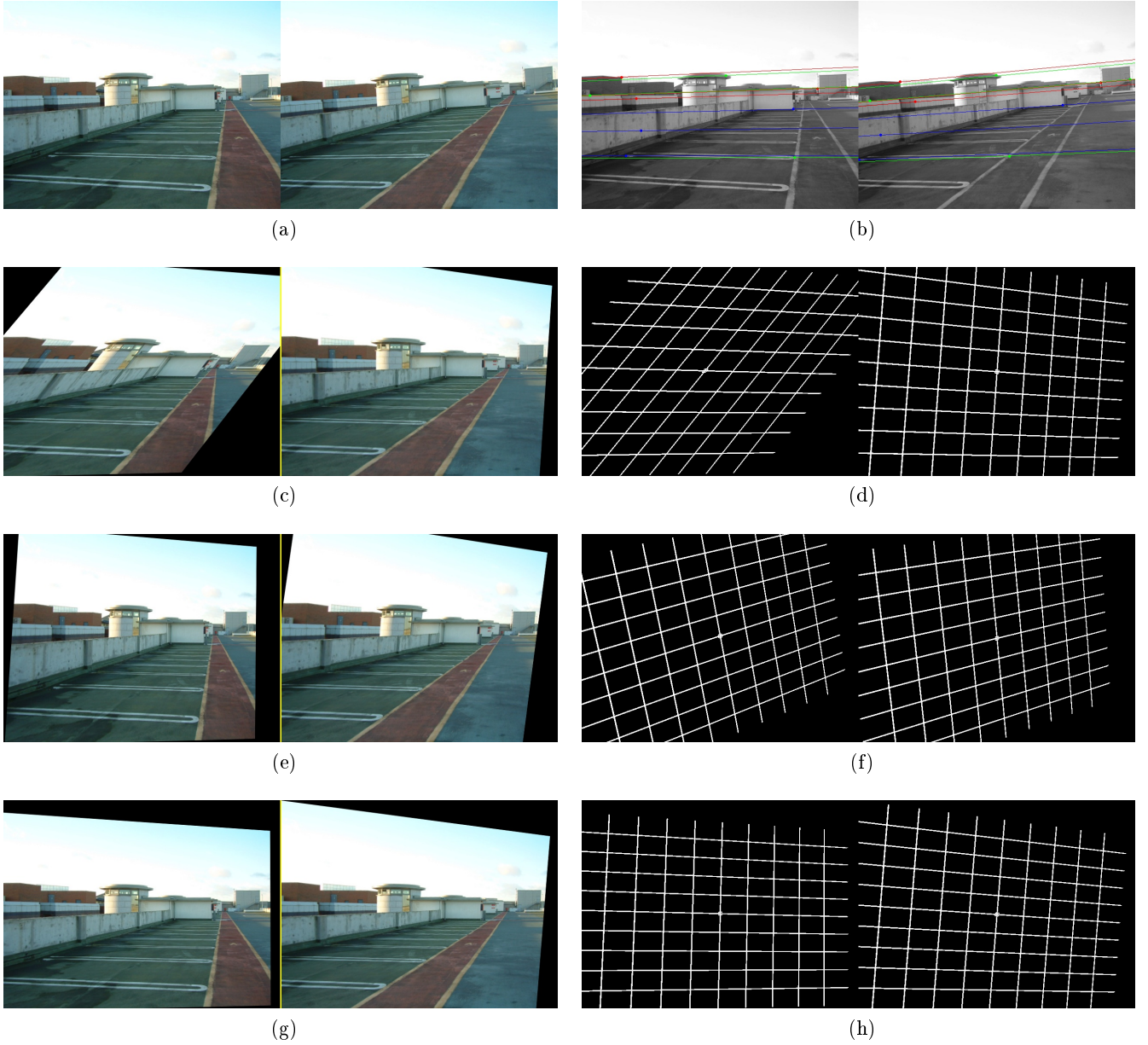


Figura A.7.4.: Rectificación de imágenes para la escena Roof, primera fila, original y original con líneas epipolares, segunda fila Hartley, tercera fila, Hartley + Zhang, cuarta fila Whelan .



## A. Apéndices

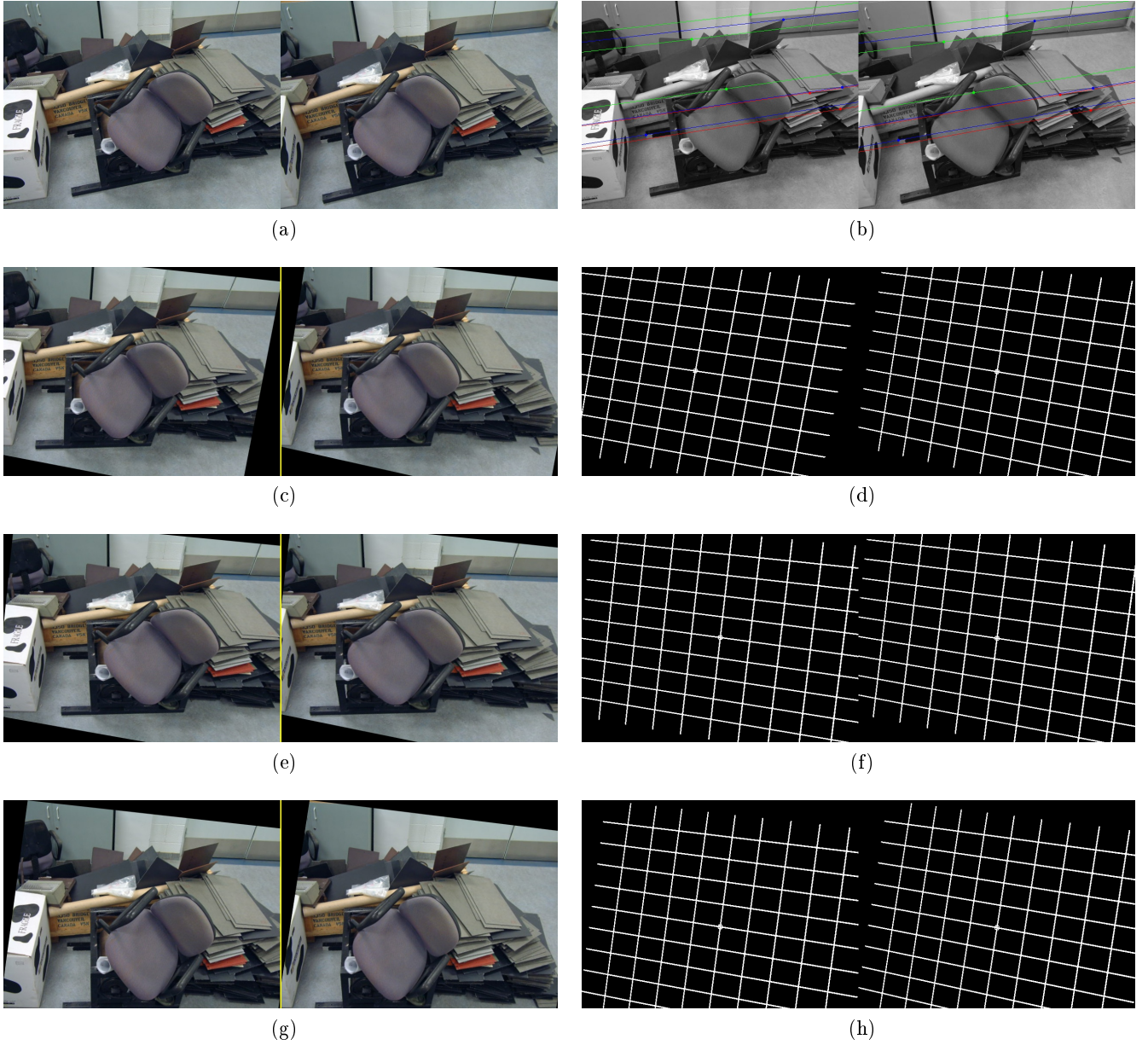


Figura A.7.5.: Rectificación de imágenes para la escena Slate, primera fila, original y original con líneas epipolares, segunda fila Hartley, tercera fila, Hartley + Zhang, cuarta fila Whelan .



## A. Apéndices

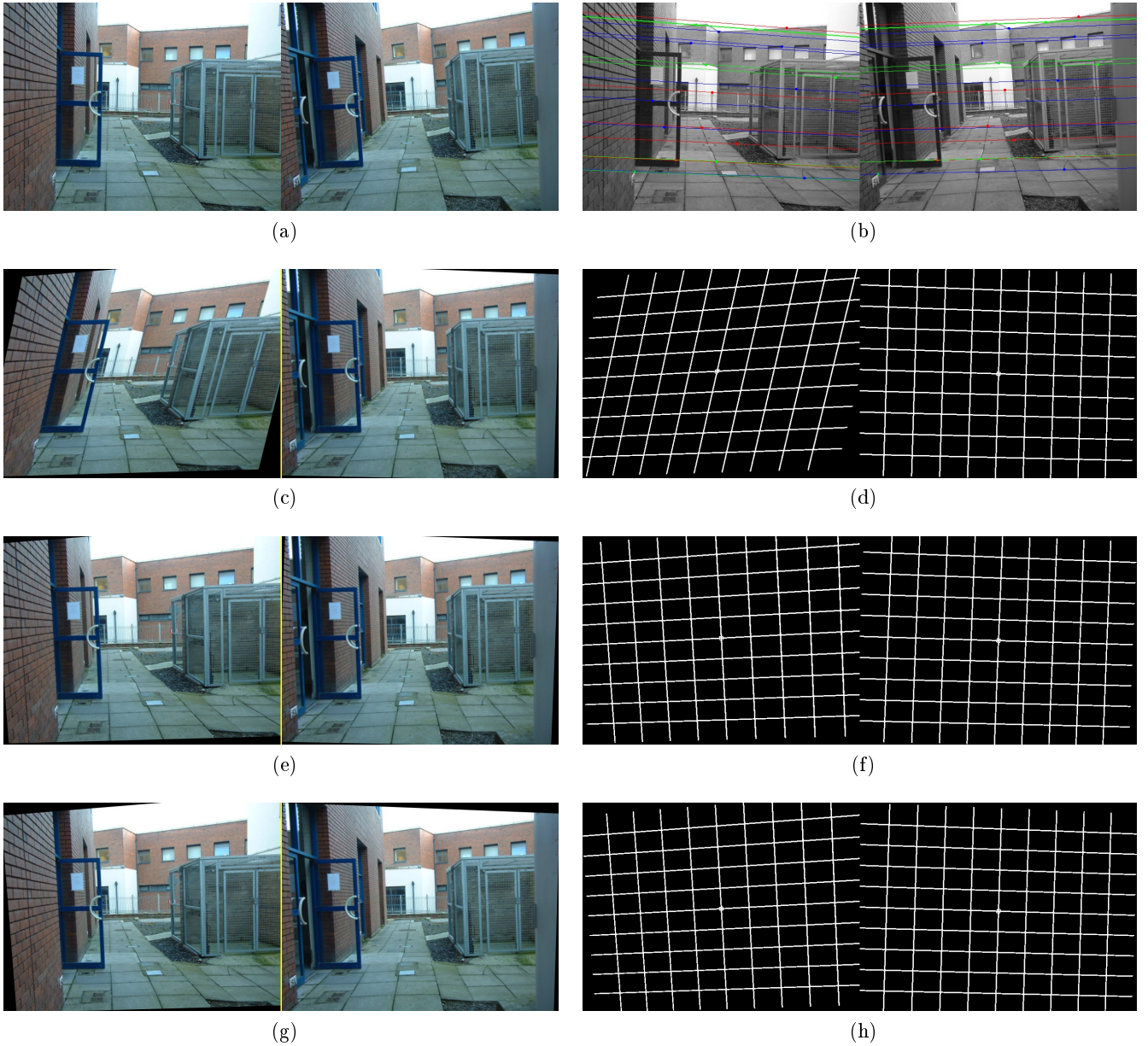


Figura A.7.6.: Rectificación de imágenes para la escena Yard, primera fila, original y original con líneas epipolares, segunda fila Hartley, tercera fila, Hartley + Zhang, cuarta fila Whelan .

# Bibliografía

- [1] Rapid Object Detection using a Boosted Cascade of Simple Features; Paul Viola y Michael Jones
- [2] A Combined Corner and Edge Detector; Cris Harris y Mike Stephens
- [3] SURF: Speeded Up Robust Features; Herbert Bay, Tinne Tuytelaars y Luc Van Gool
- [4] Speeded-Up Robust Features (SURF); Herbert Bay, Andreas Ess, Tinne Tuytelaars y Luc Van Gool
- [5] Distinctive Image Features from Scale-Invariant Keypoints; David G. Lowe
- [6] A performance evaluation of local descriptors; Krystian Mikolajczyk y Cordelia Schmid
- [7] Scale and affine invariant interest point detector; Krystian Mikolajczyk y Cordelia Schmid
- [8] Invariant Features from Interest Point Groups; Matthew Brown y David Lowe
- [9] A Flexible New Technique for Camera Calibration; Zhengyou Zhang
- [10] Close-Range Camera Calibration; D. C. Brown
- [11] Algorithms for intersecting parametric and algebraic curves I: Simple intersections; D. Manocha and J Demmel.
- [12] Determining the Epipolar Geometry and its Uncertainty; Zhengyou Zhang
- [13] múltiple View Geometry; Richard Hartley, Andrew Zissermman
- [14] In defense of the Eight-Point algorithm; Richard Hartley
- [15] A simplex method for function minimization; J. A. Nelder y R. Mead

## *Bibliografía*

- [16] Learning OpenCV; Gary Bradsky y Adrian Kaelder
- [17] Theory and Practice of Projective Rectification; Richard I. Hartley
- [18] Projective Rectification from the Fundamental Matrix; John Mallon y Paul F. Whelan
- [19] Computing Rectifying Homographies for Stereo Vision; Charles Loop y Zhengyou Zhang
- [20] Representing Shape with a Spatial Pyramid Kernel; Anna Bosh, Andre Zisserman y Xavier Muñoz
- [21] A Combined RANSAC-Hough Transform Algorithm for Fundamental Matrix Estimation; Richard den Hollander y Alan Hanjalic